

UNIVERSIDADE FEDERAL DO PARANÁ

MARTHIN SILVEIRA BORBA

ANÁLISE DA EXPRESSÃO GÊNICA EM UM MODELO DE REDE DE  
ESTABILIDADE GENÔMICA EM CANCÊR COLORRETAL E DOENÇAS  
ASSOCIADAS

CURITIBA

2016

MARTHIN SILVEIRA BORBA

ANÁLISE DA EXPRESSÃO GÊNICA EM UM MODELO DE REDE DE  
ESTABILIDADE GENÔMICA EM CANCÊR COLORRETAL E DOENÇAS  
ASSOCIADAS

Dissertação de mestrado apresentada ao  
Programa de Pós-Graduação em  
Bioinformática, Setor de Educação  
Profissional e Tecnológica, Universidade  
Federal do Paraná, área de concentração  
Bioinformática, como requisito parcial para  
a obtenção do título de Mestre em  
Bioinformática.

Orientador: Prof. Dr. Mauro A. Alves Castro

CURITIBA

2016

Catálogo na publicação  
Sistema de Bibliotecas UFPR  
Biblioteca de Educação Profissional e Tecnológica


B726	<p>Borba, Marthin Silveira</p> <p>Análise da expressão gênica em um modelo de rede de estabilidade genômica em câncer colorretal / Marthin Silveira Borba. -Curitiba, 2016. 90 p.: il., tabs, grafs.</p> <p>Orientador: Mauro Antonio Alves Castro</p> <p>Dissertação (Mestrado) – Universidade Federal do Paraná, Setor de Educação Profissional e Tecnológica, Curso de Pós-Graduação em Bioinformática.</p> <p>Inclui Bibliografia.</p> <p>1. Expressão gênica. 2. Redes reguladoras de genes. 3. Cólon (Anatomia) - Câncer. 4. Bioinformática. I. Castro, Mauro Antonio Alves. II. Título. III. Universidade Federal do Paraná.</p> <p>CDD 616.35</p>
------	---

## **TERMO DE APROVAÇÃO**

**MARTHIN SILVEIRA BORBA**

**"Análise da expressão gênica em um modelo de rede de estabilidade genômica em câncer colorretal e doenças associadas"**

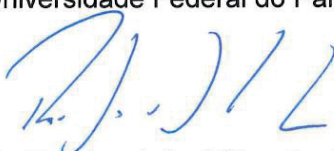
Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Bioinformática, pelo Programa de Pós-graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná, pela seguinte banca examinadora:




Dr. Mauro Antônio Alves de Castro  
Universidade Federal do Paraná - UFPR



Dr. Rodrigo Coutinho de Almeida  
Universidade Federal do Paraná - UFPR



Dr. Rodrigo Juliani Siqueira Dalmolin  
Universidade Federal do Rio Grande do Norte - UFRN



Dr. Roberto Tadeu Raittz  
Universidade Federal do Paraná - UFPR

Curitiba, 23 de fevereiro de 2017

Dedico aos meus pais e minha avó.

## **AGRADECIMENTOS**

Primeiramente agradeço à Deus, autor e idealizador da vida, da sabedoria e do conhecimento.

Ao meu orientador, Professor Dr. Mauro A. Alves Castro, muito obrigado pela oportunidade, por aceitar ser meu orientador, por toda a paciência, zelo, cuidado e direcionamento durante este período.

Professores Dra. Maria Berenice R. Steffens, Dra. Jeroniza Nunes Marchaukoski, pelas sugestões, críticas e recomendações referentes ao projeto.

Aos demais professores do programa, pelos ensinamentos e pelo conhecimento compartilhado.

Aos meus pais, muito obrigado por todo apoio, incentivo e por todo amor e carinho.

À minha querida avó, falecida durante o período do mestrado, eu agradeço por todo o carinho e amor, e dedico esse trabalho em sua saudosa e amada memória.

Aos meus familiares, que mesmo em outro estado sempre me animam, incentivam e torcem por mim.

Mariana Campioni, minha querida amada namorada e companheira. Obrigado por toda sua ajuda, paciência, apoio, incentivo, compreensão e principalmente por seu cuidado e amor.

Aos amigos do laboratório, Aniele, Sheyla, Rodrigo, Nilson e Nickolas, por toda ajuda, conversas e risadas.

Aos colegas do Grupo de Biologia de Sistemas, Vinicius, Sheyla, Kelin, Nickolas, Matheus e Clarice, por toda a contribuição e críticas que aprimoraram o projeto.

Aos demais colegas não citados mas que de alguma forma estiveram presente nessa trajetória, muito obrigado.

Ao Programa de Pós Graduação em Bioinformática e em especial à Suzana Gobetti, pelas palavras de incentivo, e por toda atenção despendida em todas as vezes que era necessário a resolução de algo referente ao programa.

A CAPES e ao Programa de Biologia Computacional, agradeço por financiar os dois anos de mestrado e a ajuda com cursos e congressos para aperfeiçoamento pessoal.

“You’re in quest for more to find the core.  
Your journey still ain’t over.  
Your quest is your purpose, go on.”

Tobias Sammet

## RESUMO

A perda da estabilidade genômica aliada a degradação das estruturas gênicas é um dos vários aspectos importantes na evolução do câncer. Os mecanismos celulares de manutenção e estabilidade do genoma são responsáveis pela integridade e pelo funcionamento correto da célula. Eles são compostos por vias genéticas que incluem *checkpoints* do ciclo celular, o reparo e recombinação do DNA, a morte celular programada (apoptose), entre outros. O câncer de colorretal é um exemplo de cancro com origem multifatorial, e surge por consequência do acúmulo de alterações genéticas e alterações somáticas, onde as células epiteliais do cólon e do reto se diferenciam em células de adenocarcinomas e posteriormente em carcinomas. O entendimento desta progressão é de substancial importância para identificar as causas da instabilidade cromossômica em células colorretais e para determinar os efeitos das diferentes formas de instabilidade genômica no comportamento biológico e clínico de tumores de cólon. O objetivo desse trabalho é observar a atividade nas vias compostas pelos genes de estabilidade e manutenção do genoma em diferentes tipos de fenótipos relacionados com o desenvolvimento do câncer colorretal. A base Ontocancro 2.0 foi utilizada para a construção de uma rede representativa de interação protéica, contento, as vias de manutenção e estabilidade do genoma. O banco de dados GEO (Gene Expression Omnibus), foi utilizado para análise de expressão diferencial da rede em amostras relacionadas tanto com tecido normal como neoplásico. A ferramenta de visualização ©Viacomplex foi utilizada para análise de expressão diferencial da rede, e o método de GSEA (Gene Set Enrichment Analysis) para confirmação estatística dos dados obtidos na análise de expressão diferencial. A análise de expressão diferencial da rede representativa das vias envolvidas no processo de estabilidade do genoma, juntamente com a análise estatística permitiu avaliar como diferentes tipos de lesões, malignas ou benignas, estão envolvidas com o desenvolvimento do câncer de colorretal.

Palavras-chave: Câncer colorretal. Estabilidade genômica. Biologia de sistemas.



## **ABSTRACT**

The loss of genomic stability coupled with of gene degradation structures is one of the several important aspects in evolution of cancer. Cellular maintenance mechanisms and stability of the genome are responsible for the integrity and correct cell functioning. These genetic pathways are composed of cell cycle checkpoints, DNA repair and recombination, programmed cell death (apoptosis), among others. Colorectal cancer is an example of cancer with a multifactorial origins, and arises from the accumulation of genetic alterations, and also by sporadic alterations, where the epithelial cells of the colon and the rectum become into adenocarcinoma cells and later carcinomas. Genetic changes resulting from loss of stability are important pathogenic and molecular steps that occur early in tumorigenesis, allowing changes in tumor suppressor genes and oncogenes, transforming cells and promoting tumor progression. Cancer progression understanding is substantial to identify the causes of chromosomal instability in colorectal cells and to determine the effects of different genomic instability forms on the biological and clinical behavior of colon tumors. The main goal was to observe the activity in the pathways composed by the genes of stability and maintenance in different fenotypes related to the development of colorectal cancer. Public data from the Ontocancro 2.0 database were used to construct a protein interaction network representing maintenance and stability pathways of the genome. Microarray data from Gene Expression Omnibus (GEO) database were used in this network model to differential expression analysis, using related samples both in normal as neoplastic tissue. The ©Viacomplex, tool for expression visualization by landscape, was used in the network differential expression analysis, and the GSEA (Gene Set Enrichment Analysis) method for statistical validation in obtained data. The network differential expression analysis of the pathways involved in genome stability and the statistical analysis allowed to evaluate how different types of lesions, malignant or benign, are involved in colorectal cancer development.

Key words: Colorectal cancer. Genomic stability. System Biology

## LISTA DE FIGURAS

FIGURA 1. DISTRIBUIÇÃO PROPORCIONAL DO TOTAL DE MORTES POR CÂNCER NO BRASIL, SEGUNDO A LOCALIZAÇÃO PRIMÁRIA DO CÂNCER, EM HOMENS, PARA OS PERÍODOS ENTRE 1999-2003 E 2009-2013. ....	13
FIGURA 2. DISTRIBUIÇÃO PROPORCIONAL DO TOTAL DE MORTES POR CÂNCER NO BRASIL, SEGUNDO A LOCALIZAÇÃO PRIMÁRIA DO CÂNCER, EM MULHERES, PARA OS PERÍODOS ENTRE 1999-2003 E 2009-2013. ....	14
FIGURA 3. ORGANIZAÇÃO DAS CRIPTAS COLÔNICAS. ....	17
FIGURA 4. FORMAÇÃO DE MASSA TUMORAL NAS CRIPTAS COLÔNICAS. ....	17
FIGURA 5. ESTÁGIOS DA PROGRESSÃO DO CÂNCER COLORRETAL. ....	18
FIGURA 6. DISTRIBUIÇÃO DAS LESÕES NA DOENÇA INFLAMATÓRIA DO INTESTINO (INFLAMMATORY BOWEL DISEASE – IBD). ....	21
FIGURA 7. CICLO DA DIVISÃO CELULAR. ....	23
FIGURA 8. REPRESENTAÇÃO DOS TRÊS CAMINHOS QUE ENGLOBAM UMA REDE PROTEICA. ....	26
FIGURA 9. REDE DE APOPTOSE E PROCESSOS ENVOLVIDOS NA ESTABILIDADE DO GENOMA. ....	28
FIGURA 10. ILUSTRAÇÃO REPRESENTATIVA DO MÉTODO DE GSEA. ....	35
FIGURA 11. GRAFO DA REDE DE ESTABILIDADE DO GENOMA. ....	38
FIGURA 12. FLUXO DE ANÁLISE FUNCIONAL, ESTATÍSTICA E GRÁFICA EM IBD. ....	42
FIGURA 13. FLUXO DE ANÁLISE FUNCIONAL, ESTATÍSTICA E GRÁFICA EM ADENOMA. ....	44
FIGURA 14. FLUXO DE ANÁLISE FUNCIONAL, ESTATÍSTICA E GRÁFICA EM CRC. ....	46

FIGURA 15. FLUXO DE ANÁLISE FUNCIONAL, ESTATÍSTICA E GRÁFICA EM MSI. ....	48
--	----

## **LISTA DE SIGLAS, TERMOS E ABREVIATURAS**

DNA - Ácido Desoxirribonucleico

mRNA - “Messenger RNA”. - RNA Mensageiro

tRNA - RNA Transportador

RNA - Ácido Ribonucleico

INCA - Instituto Nacional do Câncer

CRC - “Colorectal Cancer”. - Cancêr de Colorretal

IBD - “Inflammatory Bowel Disease”. - Doença Inflamatória Intestinal

MSI - “Microsatellite Instability”. - Instabilidade Microsatélite

MSS - “Microsatellite Stability”. - Estabilidade Microsatélite

PCD - “Programmed Cell Death”. – Morte Celular Programada

TP53 - Proteína Supressora de Tumor P53.

GS's - “Gene Sets”. - Grupo de Genes

GSEA - “Gene Set Enrichment Analysis”

GWAS - “Genome-wide Association Study”

Es - “Enrichment Score”. - Ponto de Enriquecimento

NCBI - “National Center for Biotechnology Information”.

## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO .....</b>	<b>12</b>
1.1	CARACTERIZAÇÃO E EPIDEMIOLOGIA DO CÂNCER .....	12
1.2	CÂNCER COLORRETAL .....	14
<b>2.</b>	<b>REVISÃO BIBLIOGRÁFICA .....</b>	<b>16</b>
2.1	DESENVOLVIMENTO E PROGRESSÃO ADENOMA CARCINOMA .... .....	16
2.2	INSTABILIDADE MICROSSATÉLITE (MICROSATELLITE INSTABILITY - MSI) .....	19
2.3	DOENÇA DO INTESTINO INFLAMADO (INFLAMMATORY BOWEL DISEASE - IBD) .....	20
2.4	INFORMAÇÃO GENÉTICA E ESTABILIDADE DO GENOMA .....	22
2.5	BIOLOGIA DE SISTEMAS .....	25
2.6	REDE DE GENES DE ESTABILIDADE DO GENOMA .....	27
<b>3.</b>	<b>OBJETIVOS .....</b>	<b>30</b>
3.1	OBJETIVOS GERAIS .....	30
3.2	OBJETIVOS ESPECÍFICOS .....	30
<b>4.</b>	<b>MATERIAS E MÉTODOS .....</b>	<b>31</b>
4.1	OBTENÇÃO DE DADOS PARA CONSTRUÇÃO DA REDE .....	31
4.2	RECUPERAÇÃO DA INTERAÇÃO ENTRE PROTEÍNAS .....	32
4.3	CONSTRUÇÃO DA REDE .....	33
4.4	OBTENÇÃO E ANÁLISE DOS DADOS DE EXPRESSÃO GÊNICA .....	33
4.5	ANÁLISE DE ENRIQUECIMENTO DE GRUPO DE GENES .....	34
4.6	VISUALIZAÇÃO GRÁFICA DA ATIVIDADE FUNCIONAL DA REDE DE ESTABILIDADE DO GENOMA .....	35
<b>5.</b>	<b>RESULTADOS .....</b>	<b>37</b>
5.1	REDE DE ESTABILIDADE DO GENOMA .....	37
5.2	ANÁLISE DE EXPRESSÃO GLOBAL .....	40
5.2.1	DOENÇA INFLAMATÓRIA INTESTINAL (IBD) .....	40
5.2.2	ADENOMA .....	43
5.2.3	CANCER DE COLORRETAL (CRC) .....	45
5.2.4	INSTABILIDADE MICROSSATÉLITE (MSI) .....	47

<b>6.</b>	<b>DISCUSSÃO .....</b>	<b>50</b>
6.1	CONSIDERAÇÕES GERAIS .....	50
6.2	CONSIDERAÇÕES ESPECÍFICAS .....	52
<b>7.</b>	<b>CONCLUSÃO .....</b>	<b>55</b>
<b>8.</b>	<b>PERSPECTIVAS .....</b>	<b>57</b>
<b>9.</b>	<b>REFERÊNCIAS .....</b>	<b>58</b>
	<b>ANEXO - LISTA DE GENES RETIRADOS DO ONTOCANCRO 2.0</b>	
	<b>PARA CONFEÇÃO DA REDE DOS PROCESSOS DE</b>	
	<b>ESTABILIDADE DO GENOMA .....</b>	<b>65</b>

## 1. INTRODUÇÃO

### 1.1 CARACTERIZAÇÃO E EPIDEMIOLOGIA DO CÂNCER

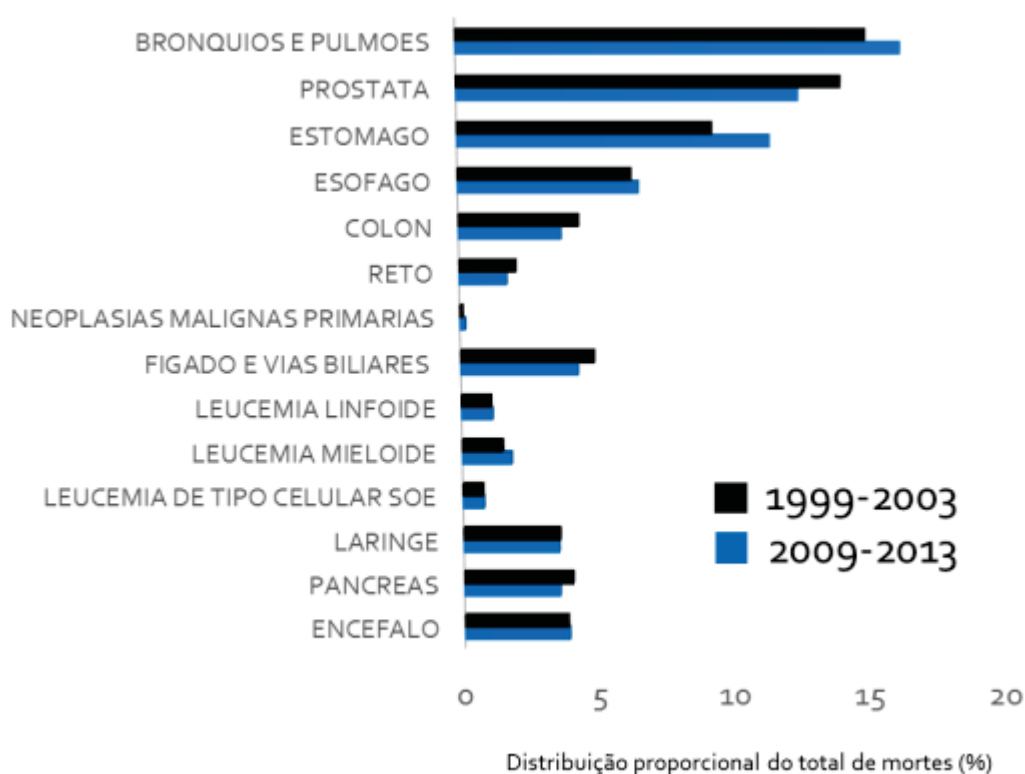
Pode-se definir como câncer a formação de um tumor maligno caracterizado pelo crescimento desordenado de células anormais, que como consequência acarreta em invasão de tecidos e órgãos adjacentes. Sua disseminação está intimamente ligada à progressão da doença, um processo denominado metástase (INCA, 2015). O câncer geralmente está dividido em dois grupos distintos, os de origem hereditária e os que são resultado de lesões cumulativas (câncer somático). Devido a diversidade de tipos celulares e o grande número de fatores que o originam, podemos considerar o câncer como uma doença multifatorial, onde a sua estabilidade e a instabilidade estão diretamente ligados a sua epidemiologia.

A progressão do câncer, também classificado como neoplasia maligna, é um processo complexo, envolvendo uma variedade de rotas metabólicas e grupos distintos de células, além de evoluir em diferentes escalas de tempo e em diferentes locais, onde vários fatores contribuem para a doença (Anderson; Quaranta, 2008). Considerando isso, o termo “câncer” está ligado a conjuntos altamente heterogêneos de doenças que podem se desenvolver em qualquer parte do corpo, mas que possuem como características comuns, o crescimento celular desordenado, a invasão de tecidos adjacentes ao sítio em que se desenvolveu e a formação de metástases (Kufe et al., 2005).

Historicamente as neoplasias de origem hematológicas têm sido classificadas e tratadas separadamente das demais neoplasias, as quais são caracterizadas como neoplasias sólidas ou tumores sólidos (Bennett et al., 1976; Harris et al., 1995; Harris et al., 1999; ICD-O-1, 1976; Rappaport et al., 1971). Durante um longo período nos estágios iniciais da progressão tumoral as neoplasias sólidas geram massas tumorais que geralmente estão limitadas por barreiras teciduais, até quando adquirem uma forma invasiva e conseguem ultrapassar os limites teciduais do local de origem e invadem novos sítios (Jones et al., 2008). No decorrer da progressão do tumor, as neoplasias sólidas desenvolvem inúmeras alterações citogenéticas e moleculares (Hoglund et al., 2002b), e nos estágios mais avançados de crescimento tumoral elas expressam grande heterogeneidade intratumoral, o que ocasiona a

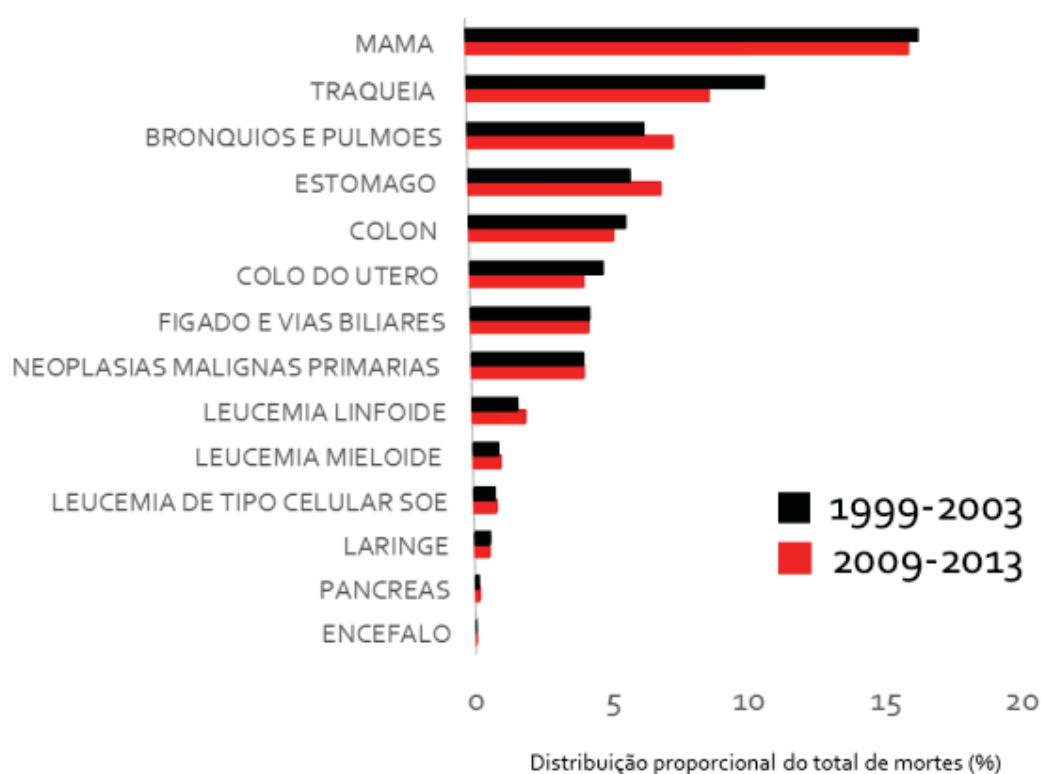
formação de múltiplos clones aberrantes (Albertson et al., 2003; Backvall et al., 2005; Glockner et al., 2002; Harada et al., 1998; Heng et al., 2006; Losi et al., 2005; Mitelman, 2000; Mitelman et al., 1997; Orndal et al., 1993).

No Brasil a maior incidência de câncer está ligada ao desenvolvimento de neoplasias sólidas (INCA, 2016), enquanto as neoplasias de origem hematológica, como linfomas e leucemias, registram índices menores de ocorrência.



**Figura 1.** Distribuição proporcional do total de mortes por câncer no Brasil, segundo a localização primária do câncer, em Homens, para os períodos entre 1999-2003 e 2009-2013.  
 FONTE: INCA/BR (Atlas de Mortalidade por Câncer, 2016). SOE: Sem Outras Especificações.





**Figura 2.** Distribuição proporcional do total de mortes por câncer no Brasil, segundo a localização primária do câncer, em Mulheres, para os períodos entre 1999-2003 e 2009-2013. FONTE: INCA/BR (Atlas de Mortalidade por Câncer, 2016). SOE: Sem Outras Especificações.

Os gráficos das **Figuras 1 e 2** mostram que o maior número de mortes por câncer em ambos os sexos está diretamente ligado ao desenvolvimento de neoplasias solidas. O câncer sempre foi uma doença relevante e de grande interesse de estudos científicos em decorrência dos elevados índices de incidência e mortalidade. Os esforços no sentido de elucidar processos e mecanismos ainda desconhecidos da doença são de fundamental importância para propor novas formas de tratamento e prevenção.

## 1.2 CÂNCER COLORRETAL

O câncer colorretal (CCR), ou em inglês *colorectal cancer* (CRC) é comum em todo o mundo e continua sendo a segunda causa de morte relacionada ao câncer nos países ocidentais (Siegel et al., 2011). Apesar das recentes melhorias nas estratégias de rastreio e no desenvolvimento de tratamentos mais eficazes para a CRC, o prognóstico de CRC ainda é fraco.

As mutações genéticas são há muito tempo conhecidas por serem importantes na formação do câncer. No entanto, sabe-se que fatores externos também são reconhecidos como contribuintes significativos para o desenvolvimento do câncer recentemente (Waddington 1942; Bird et al., 2002).

O desenvolvimento do câncer colorretal está diretamente envolvido com o acúmulo de aberrações e erros moleculares, tanto estruturais como falhas na comunicação e interação entre produtos funcionais de genes responsáveis por processos e sinalização dos mecanismos celulares. Mutações nos genes *KRAS*, *HRAS* e *NRAS* da família de oncogenes *RAS*, são conhecidas como um fator de risco significativo para a progressão e desenvolvimento da malignidade da doença. Os passos seguintes na evolução de CRC metastático são mutações estendidas ao gene *BRAF*, o que altera a estabilidade à níveis cromossomais, bem como mutações no genes *PIK3CA*, geralmente acompanhadas pelas mudanças causadas nos genes da família *RAS* (Chen et al., 2016). As causas moleculares envolvidas com a progressão da doença estão fundamentadas pelo modelo de progressão adenoma-carcinoma. A medida que eventos moleculares promovem mutações, conjunto de genes adquirem maior probabilidade de sofrer erros ao longo desses estágios de desenvolvimento (Waddington 1942; Bird et al., 2002).

A estabilidade do sistema de identificação e dos mecanismos de reparo do DNA afetados por mutações a níveis moleculares causam mudanças conformacionais responsáveis pela aquisição de um perfil de instabilidade, a níveis celulares, comprometendo os sistemas de reparo e estabilidade da célula. Ainda se tratando de instabilidade, existe alguns tipos de câncer, que assumem um comportamento instável decorrente de mutações, duplicações e deleções em regiões específicas do DNA. Essas mudanças geram um comportamento totalmente instável em células tumorais, dificultando a identificação de um padrão em casos de cânceres instáveis.

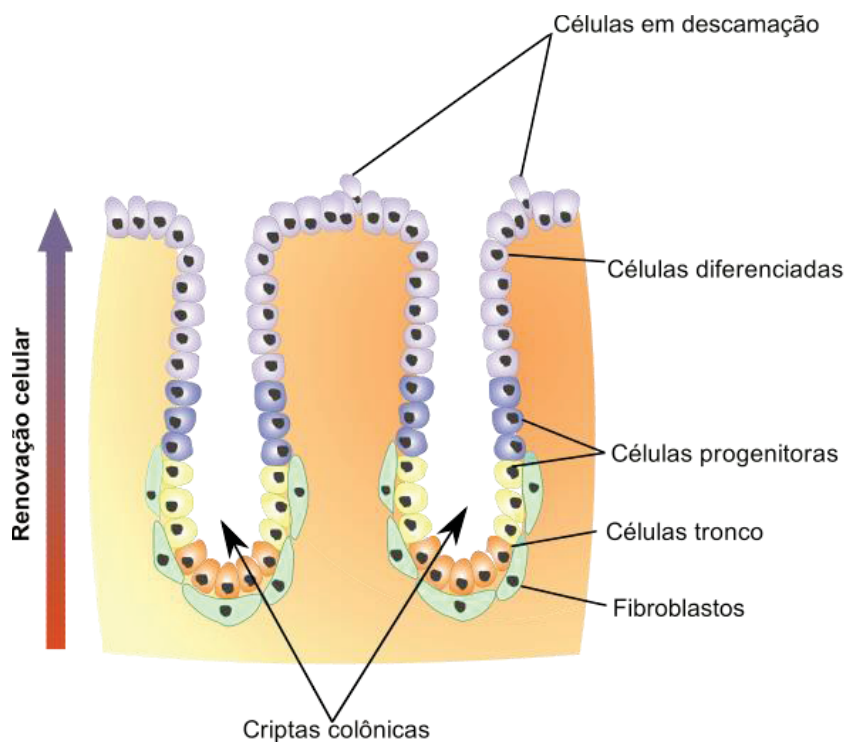
A biologia de sistemas possibilita o cruzamento de dados gerados em laboratório, e sua instrumentalidade auxilia nas interpretações e na tomadas de decisões sobre a progressão e os aspectos funcionais no câncer.

## 2. REVISÃO BIBLIOGRÁFICA

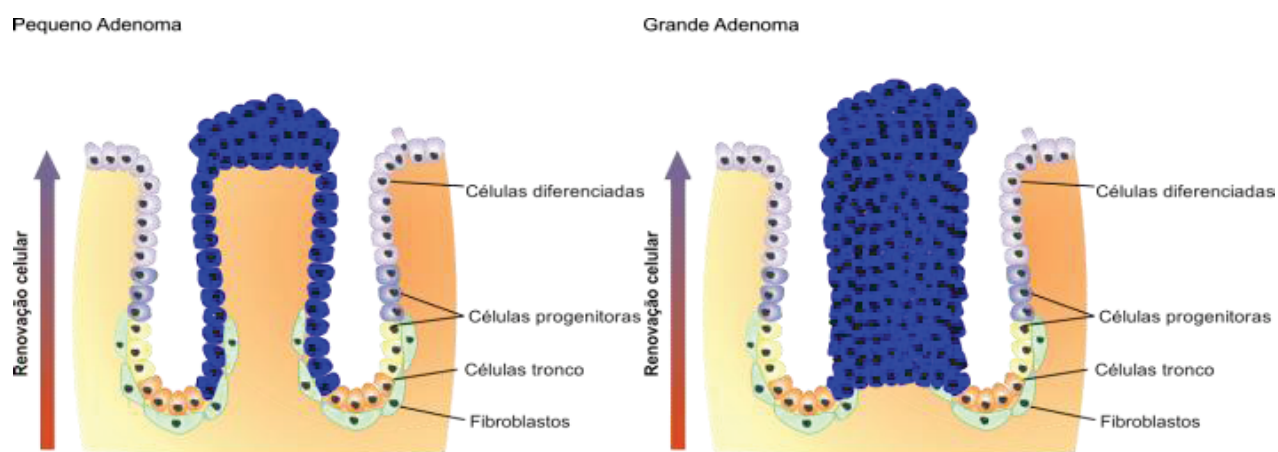
### 2.1 DESENVOLVIMENTO E PROGRESSÃO ADENOMA CARCINOMA

O câncer de origem colorretal possui carcinogênese bem descrita, sendo um modelo para o estudo de progressão tumoral (Hahn & Weinberg, 2002; Kufe et al., 2005; Castro, 2009). A progressão do câncer colorretal está fundamentada no modelo adenoma-carcinoma, postulada por Fearon & Vogelstein (Fearon & Vogelstein, 1990), e baseia-se em uma sequência de eventos moleculares, em que cada estágio exhibe um conjunto de genes com maior probabilidade de mutação, sendo que o processo tem início em um tecido normal. Assim, o entendimento da morfologia do tecido normal é importante para descrever a progressão.

A **Figura 3** ilustra a organização das criptas colônicas de um tecido sadio. Em condições normais, as células basais migram para a superfície da cripta mantendo um fluxo contínuo de divisão celular e, à partir da base, ocorre a substituição das células do ápice, as quais sofrem um processo de esfoliação pelo funcionamento dinâmico do intestino (Humphries & Wright, 2008). O processo de renovação leva de três a seis dias e quando as taxas de mitose são mais elevadas que as taxas de perda celular então ocorre a formação de uma massa tumoral (Fodde et al., 2001) (**Figura 4**).



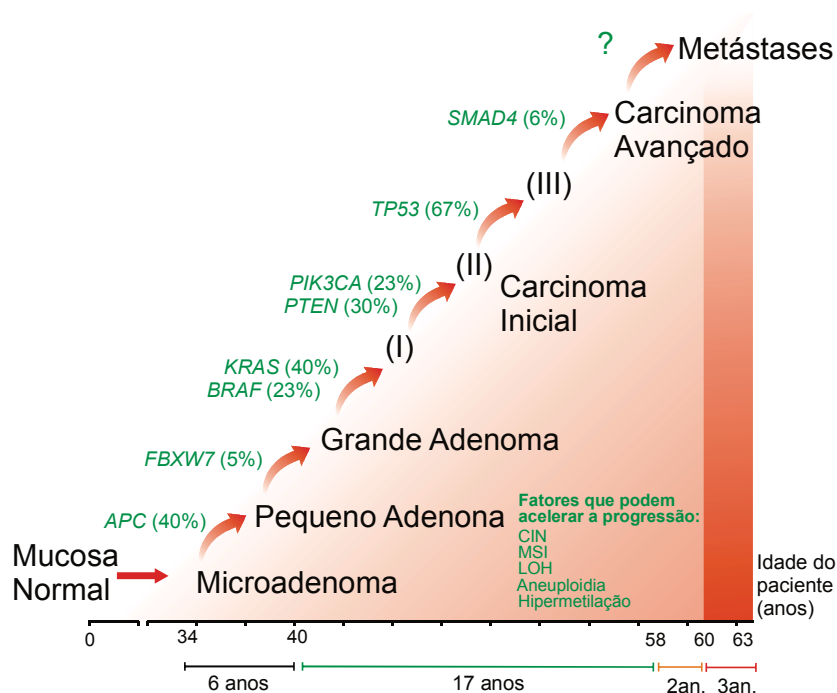
**Figura 3.** Organização das criptas colônicas (adaptado de Castro, 2009).



**Figura 4.** Formação de massa tumoral nas criptas colônicas. Processo de eventos hiperplásicos e displásicos responsáveis pela formação de um carcinoma de colorretal (adaptado de Castro, 2009).

No aspecto histológico, o primeiro estágio da progressão tumoral é caracterizado pelo acúmulo de células na superfície da cripta, chamado de processo hiperplásico ou hiperplasia, seguido por modificações na organização tecidual, denominado displasia. Com esse processo já instaurado, focos do tecido aberrante projetam-se para fora da superfície da mucosa normal em direção à luz do tubo digestivo, originando pólipos (Castro, 2009). Pólipos displásicos já são considerados lesões pré-malignas, classificados como adenomas, e podem evoluir pra neoplasias malignas (Cotran et al., 2000).

Já em relação ao aspecto molecular, o primeiro estágio da progressão tumoral parece estar relacionado a mutação do gene *APC*, o qual está alterado em cerca de 40% dos adenomas e carcinomas de cólon e reto (**Figura 5**).



**Figura 5.** Estágios da progressão do câncer colorretal CIN: instabilidade ao nível cromossomal; MSI: instabilidade microssatélite; LOH: perda de heterozigossidade. Adaptado de Jones et al., 2008). Percentuais indicam fração de tumores com genes mutados ao final da progressão, segundo o projeto COSMIC onde foram avaliados 117 tumores e no estágio de adenoma o gene *APC* está mutado em 42% dos casos (Forbes et al., 2008). FONTE: CASTRO M.A.A., 2009.

No modelo de progressão ilustrado na **Figura 5**, o gene *APC* atua como inibidor da atividade da proteína  $\beta$ -catenina, a qual tem a capacidade de induzir a expressão de genes envolvidos em processos de adesão e divisão celular, dentre

outros. A inativação da via APC/ $\beta$ -catenina estaria associada à progressão tumoral até o estágio de microadenoma (Ceol et al., 2007; Fodde et al., 2001). Mutações na família RAS de oncogenes (*i.e.* genes *KRAS*, *HRAS* e *NRAS*) e ativação da via KRAS/BRAF seriam os próximos eventos do ponto de vista genético (Kinzler & Vogelstein, 1996), e a aquisição de um fenótipo instável é responsável por acelerar a progressão (*e. g.* instabilidade ao nível dos cromossomos) (Jefford et al., 2006). Os próximos passos envolveriam mutações em genes que controlam a expressão do gene *TP53*, bem como alterações em vias responsáveis pela transição de tumores benignos (adenoma) para tumores malignos (carcinomas) (Jones et al., 2008).

O arranjo topográfico da proliferação celular da mucosa colônica facilita a obtenção de amostras em diferentes fases de desenvolvimento, e isso evidencia o fato do câncer de colorretal ter se tornado um modelo de estudo de progressão tumoral (Castro, 2009)<sup>2</sup>.

## 2.2 INSTABILIDADE MICROSSATÉLITE (MICROSATELLITE INSTABILITY - MSI)

As repetições satélites do DNA podem ser classificadas de acordo com a extensão da sequência repetitiva em satélites, minissatélites e microssatélites. Regiões microssatélites, também conhecidas como repetições em “tandem” ou “short tandem repeats” (STR), encontram-se distribuídas aleatoriamente por todo o genoma humano, podendo ocorrer a cada 50.000 a 100.000 pares de bases aproximadamente. Embora estas sequências apresentem características extremamente polimórficas na população, elas são herdadas e conservadas de uma geração para a outra, e podem ser encontradas tanto em tecidos normais quanto tumorais, e assim serem utilizadas como marcadores polimórficos. O seu tamanho varia de um a seis nucleotídeos, que se repetem em poucas dezenas (Weber & May, 1989). As repetições mais comuns são (A)<sub>n</sub>, (T)<sub>n</sub>, (CA)<sub>n</sub>, (GT)<sub>n</sub>, mas também aparecem como trinucleotídeos ou como tetranucleotídeos (Marra & Boland, 1996; Debrauwere et al., 1997).

As alterações encontradas nessas regiões, envolvendo deleção ou inserção dessas pequenas unidades repetitivas, caracterizam a instabilidade microssatélite. Segundo o projeto ENCODE apenas 1% do genoma codifica proteínas, porém as sequências microssatélites não ocorrem nessa porção do genoma. Sabe-se que elas podem atuar

diretamente na regulação do gene ou indiretamente com ponto de instabilidade no processo de recombinação (Saletti et al., 2001).

Estudos feitos em *Escherichia coli* e *Scacaromices cerevisiae* revelaram deficiências no sistema “Mismatch Repair” (MMR) que alteram essas sequências, resultando na instabilidade instabilidade microssatélite (Levinson et al., 1987). Estudos realizados a partir de 1993 mostraram que essas regiões exibiam alterações de tamanho em câncer de colorretal (Strand et al., 1993).

A observação de um grande número de alterações nas sequências de microssatélites em um determinado tecido tumoral demonstra a ausência de uma função normal de reparo de DNA e isso representa consequentemente, uma evidência indireta de que existe uma deficiência na ação das proteínas de reparo causada pela presença de mutações. A este tipo de defeito genético denominamos erros de replicação em RER (Recombinational Repair - RER) (Lam et al., 2014).

A MSI é uma marca registrada encontrada em alguns cancros de colorretal. Oitenta e cinco a 90% dos CCRs hereditários (HPC) apresentam MSI, enquanto ocorre apenas em 10 – 15% dos CCRs esporádicos (AALTONEN et al., 1993; IONOV et al., 1993; THIBODEAU et al., 1993), porém a literatura ressalta que o câncer colorretal associado a erros de replicação irá apresentar o mesmo tipo de localização e comportamento biológico independentemente de sua natureza esporádica ou hereditária.

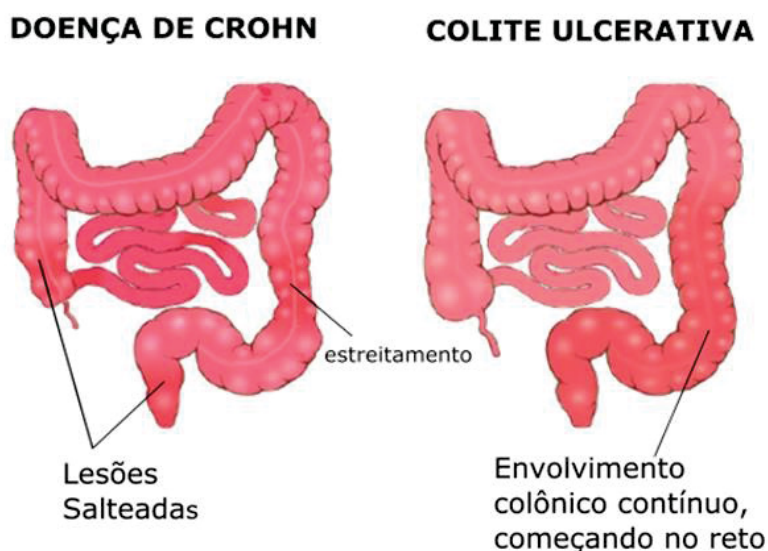
### 2.3 DOENÇA DO INTESTINO INFLAMADO (INFLAMMATORY BOWEL DISEASE - IBD)

A doença do intestino inflamado (IBD), também chamada de doença inflamatória intestinal, é uma condição crônica resultante da ativação imune inadequada da mucosa e pode ser dividida em duas formas principais, a doença de Crohn e a colite ulcerativa. A distinção entra ambas é baseado nas diferenças de distribuição de locais afetados e na expressão morfológica da doença (Kumar et al., 2013). Enquanto a colite ulcerativa se limita a inflamação do colo, a doença de Crohn atinge qualquer parte da região gastrointestinal (Khor et al., 2016) (**Figura 6**).

Por possuir um padrão inflamatório característico, e a doença estar ligada ao desenvolvimento do câncer colorretal, IBD é um excelente objeto para identificação



de alterações de expressão gênica em um tecido acometido por lesões inflamatórias em relação a um tecido sadio de colo e reto.



**Figura 6.** Distribuição das lesões na doença inflamatória do intestino (Inflammatory Bowel Disease – IBD). Distinção morfológica entre a doença de Crohn e colite ulcerativa (adaptado de Lodish, 2003).

A análise de genes e locus genéticos envolvidos na IBD mostram que várias vias são cruciais para a homeostase intestinal, incluindo a função de barreira, restituição epitelial, defesa microbiana, regulação imune inata, espécies reativas de oxigênio, autofagia, regulação da imunidade adaptativa, stress endoplasmático reticular e vias metabólicas associadas com a homeostase celular (Khor et al., 2016). Os primeiros estudos sugerem a existência de alelos de proteção e predisposição. Vias biológicas são ainda mais relevantes na doença quando vários componentes estão agindo em conjunto como fatores de risco (Momozawa et al., 2010).

Do ponto de vista molecular, o gene *NOD2* está descrito como gene de suscetibilidade na doença de Crohn. Ele codifica uma proteína que se liga a peptídeoglicanos bacterianos intracelulares e subsequentemente ativa NF- $\kappa$ B (fator nuclear kappa B). Variantes de *NOD2* associados à doença são menos eficazes no combate e reconhecimento de microrganismos luminiais, que entram na lamina própria e assim causam reações inflamatórias. Outros dados sugerem que *NOD2* pode regular a resposta imune para evitar excesso de ativação de microrganismos luminiais (Kumar et al., 2013).



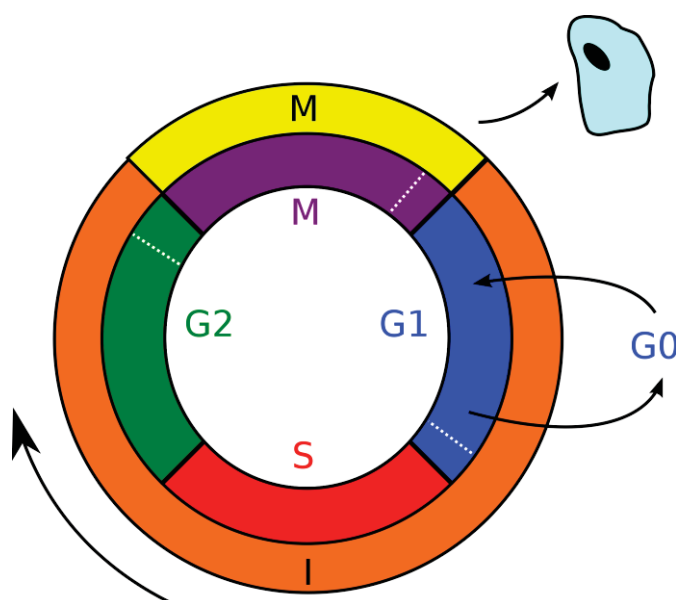
Estudos de associação genômica (Genome-wide Association Study - GWAS) que avaliam polimorfismos de nucleotídeo único têm sido utilizados para ampliar a busca por genes associados a IBD. Sequenciamentos em IBD recuperam variantes já conhecidas, e novas variantes de *CARD9*, *NOD2* e *IL23R*, com efeitos independentes sobre o risco de doença (Momozawa et al., 2010). O consórcio IBDGenetics encontrou 99 locus gênicos definitivos envolvidos com a suscetibilidade a IBD. O mesmo estudo revelou que outros 71 estão relacionados à doença de Crohn e outros 47 relacionados à colite ulcerativa, representando aproximadamente 23 e 16% da herdabilidade da doença respectivamente. Além dos exemplos mostrados acima, existem outros genes identificados por GWAS, porém dois genes relacionados a doença de Crohn possuem papel de destaque juntamente com *NOD2*, eles são *ATG16L1* e *IRGM*, relacionados à autofagia, e em processos imunes de eliminação de bactérias intracelulares (Kumar et al., 2013; Khor et al., 2016).

O câncer colorretal está associado como uma das principais complicações a longo prazo da colite ulcerativa e da doença de Crohn, onde o processo de displasia apresentado no modelo Fearon & Vogelstein atua como passo importante para a diferenciação celular de adenomas em carcinomas. Pelo fato de adenomas de colón ocorrerem em pessoas com IBD, o diagnóstico tardio mascara a real condição da progressão da doença (Kumar et al., 2013).

## 2.4 INFORMAÇÃO GENÉTICA E ESTABILIDADE DO GENOMA

Genomas são codificados em sequências formadas pelo ácido desoxirribonucleico (DNA). O DNA é um composto orgânico que possui as informações genéticas responsáveis para funcionamento e do desenvolvimento dos seres vivos. As células de um organismo possuem o mesmo genoma, independentemente da fase de desenvolvimento e condições impostas pelo ambiente (Keedwell et al., 2005). A célula codifica diferentes partes do genoma em proteínas e a informação genética está contida em forma de código por uma sequência específica formada a partir de quatro diferentes tipos de nucleotídeos. As informações contidas no genoma de uma célula serão transferidas as células filhas através do processo de divisão celular, a sequência de eventos necessários para o

que aconteça a divisão celular são chamados de ciclo celular, divididos em quatro fases principais:  $G_1$ , S,  $G_2$  e M, como mostra a **Figura 7**. Na fase  $G_1$  também conhecida como fase de crescimento, ocorre uma intensa síntese de moléculas distintas importantes para a sobrevivência celular, como RNA, proteínas estruturais e enzimas. Na fase S acontece a duplicação do DNA, seguida pela fase  $G_2$  onde a síntese de moléculas e organelas relacionadas ao processo de divisão celular é intensa. O ciclo celular é concluído na fase M, quando ocorre a segregação cromossômica (Watson et al., 2014).



**Figura 7.** Ciclo da divisão celular: o ciclo se inicia na fase  $G_1$  ou  $G_0$  com a síntese de moléculas, passa para a fase S onde há duplicação do DNA e segue pela fase  $G_2$ , e é concluído na fase M, na segregação cromossômica. FONTE: Wikiciencias: Ciclo\_Celular, 2015 (conteúdo disponibilizado nos termos da Creative Commons, CC BY-NC-AS 3.0).

Também conhecidas como fases de intervalo, as fases  $G_1$  e  $G_2$  do ciclo celular, possuem dois propósitos. Primeiro fornecem tempo para a célula se preparar para a fase seguinte e para verificar se a fase anterior foi devidamente completada. Nessas duas fases ocorrem pontos de verificação do DNA (checkpoints), onde as células com o DNA danificado param o ciclo em  $G_1$ , antes da síntese, ou em  $G_2$  antes da mitose (fase M), para impedir que os processos do ciclo celular continuem com cromossomos danificados. Estes processos são de extrema importância para

obter-se elementos funcionais coesivos com o código original. Entre as fases  $G_1$  e  $G_2$ , a fase S ou fase de síntese ocorre o processo de duplicação do DNA.

As fases do ciclo celular englobam acontecimentos estruturais dinâmicos, diretamente envolvidos com a funcionalidade de mecanismos celulares, como por exemplo, a transcrição e a tradução entre outros. A transcrição é a síntese de uma molécula de RNA mensageiro à partir de uma sequência biológica codificada e contida na molécula de DNA. O RNA mensageiro contém a informação que determina a sequência de aminoácidos de uma proteína, sendo assim, é responsável por direcionar a síntese proteica. Em um fragmento de uma molécula de DNA, mais precisamente em um gene delimitado por uma região promotora (local onde tem início a síntese, processada da extremidade 5' para a 3' na cadeia de DNA). A RNA polimerase sintetiza o RNA percorrendo a cadeia de DNA até a outra extremidade do gene onde encontra uma região terminal. Fatores de transcrição são proteínas que ligam-se ao DNA e auxiliam a enzima RNA polimerase a romper as ligações de hidrogênio entre as bases nitrogenadas presentes no DNA. Ao final do processo de transcrição, a RNA polimerase se dissocia e assim se obtém uma nova molécula de RNA mensageiro que será utilizada no processo de tradução de proteínas. Por sua vez, a tradução é um processo de síntese de proteínas por meio de uma molécula de RNA mensageiro, dividida em três partes. Iniciação, prolongação e finalização. A molécula que contém o anticódon complementar ao códon presente no RNA mensageiro, o RNA transportador, liga-se ao RNA mensageiro e por meio de uma série de eventos no ribossomo, como o reconhecimento e pareamento de cada códon complementar é formada uma proteína, por adição de aminoácidos correspondentes aos códons (Keedwell et al., 2005).

O sucesso desses processos, é importante para todas as funções do ciclo e funcionamento celular. Erros estruturais, de interpretação ou de sinais, levam a uma sequência de acontecimentos onde a informação é comprometida e tem efeitos sobre a funcionalidade de todo o sistema. Em outras palavras estes erros são responsáveis pela cascata de eventos subsequentes que causam mudanças em um ciclo celular tão bem definido, alterando o resultado final e o funcionamento correto da célula.

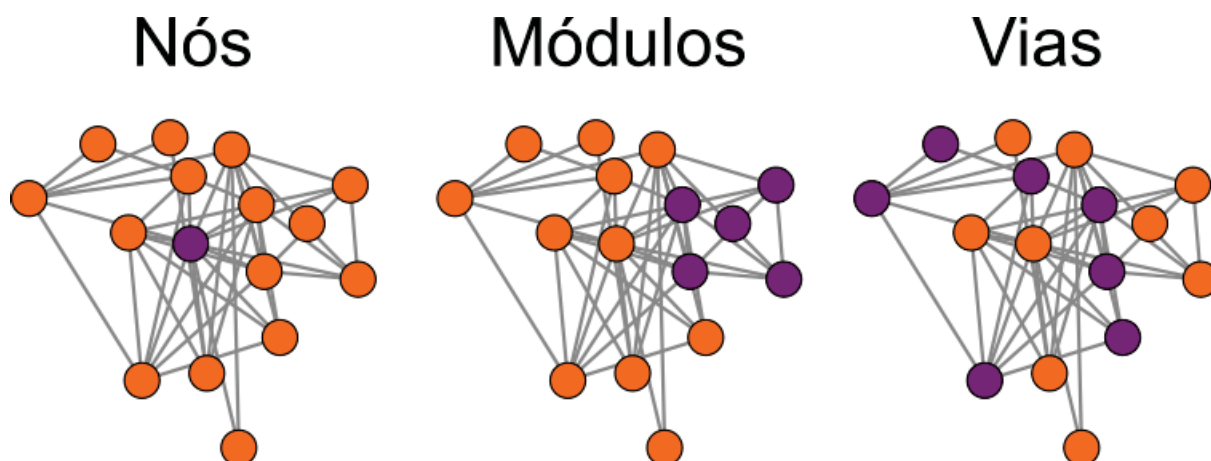
A perda da estabilidade genômica associada com a deterioração genética causada por mutação é um dos muitos aspectos importantes na evolução de carcinomas. Mutações pontuais em genes de reparo fazem com que esses mecanismos fiquem deficientes e percam sua função de identificação e correção em regiões acometidas por outros erros. Os mecanismos de manutenção são responsáveis pela correção e estabilidade do genoma, garantindo a integridade do DNA e a sobrevivência da célula.

## 2.5 BIOLOGIA DE SISTEMAS

Avanços científicos nas áreas de biologia molecular, da bioinformática e o desenvolvimento de novas tecnologias com maior processamento de dados e de menor custo, tornaram possível a obtenção de informações mais completas e de maior capacidade para definir abordagens mais detalhadas. Dessa forma é possível explicar os fenômenos biológicos não por meio da identificação e estudo detalhado de cada componente funcional, mas também pela construção de redes que envolvem suas ligações e interações, desvendando assim propriedades antes não conhecidas, e mostrando um novo panorama para o sistema estudado.

Os diagramas de interação constituem uma maneira interativa de analisar um determinado processo biológico a níveis moleculares (Barabasi et al., 2004). Estes diagramas, também denominados redes gênicas ou proteicas, representam o conjunto das interações moleculares, seja ele relativo a um determinado processo celular, ou a todo repertório de genes/proteínas do organismo (interatoma).

Pode-se chamar uma rede de interações de grafo. A construção de grafos representativos para associação de proteínas é definida por três elementos principais: nós, módulos e vias (**Figura 8**). Os nós abrangem as reações que estão associadas com uma determinada proteína. A sinalização de nós conectados indicam ativação de proteínas que exercem funções importantes dentro do complexo funcional da célula (Papin et al., 2005).



**Figura 8.** Representação dos três caminhos que englobam uma rede proteica: nós, módulos de sinalização e vias com formação linear. Em laranja está demonstrada a representação de uma rede. Na primeira imagem em roxo está destacado um exemplo de nó, na segunda imagem um exemplo de módulo, e na terceira imagem a representação de uma via que conecta duas partes da rede (adaptado de Papin et al., 2005).

Os módulos são constituídos por grupos de proteínas que funcionam juntas sobre certas condições. Os módulos representam a função relacionada entre proteínas, ou seja, proteínas altamente conectadas que não atuam nas mesmas vias somente (Papin et al., 2005).

As vias são elementos que conectam as entradas de sinalizações e as sinalizações de saída, ou seja, uma determinada via pode ser a delimitação de todas as ligações de um fator de crescimento ao seu receptor até a próxima ativação de um fator de transcrição que induz a expressão de genes alvos (Papin et al., 2005). As vias determinam funcionalmente uma ligação entre proteínas que apresentam funções equivalentes. Se a função de uma dessas proteínas é conhecida, então pode-se inferir que essas proteínas agem na mesma via (Eisenberg et al. 2000).

Com o advento da biologia de sistemas, foram desenvolvidas abordagens integrativas capazes de analisar informações complexas como é o caso das análises de expressão gênica em larga escala por microarranjos (microarrays), podendo retirar informações em diferentes níveis, abrangendo desde o perfil global da transcrição gênica celular (Rybarczyk-Filho et al., 2011), até alterações em grupos de genes correlacionados (Subramanian et al., 2005) a determinados processos biológicos (ViaComplex Castro et al., 2009).

## 2.6 REDE DE GENES DE ESTABILIDADE DO GENOMA

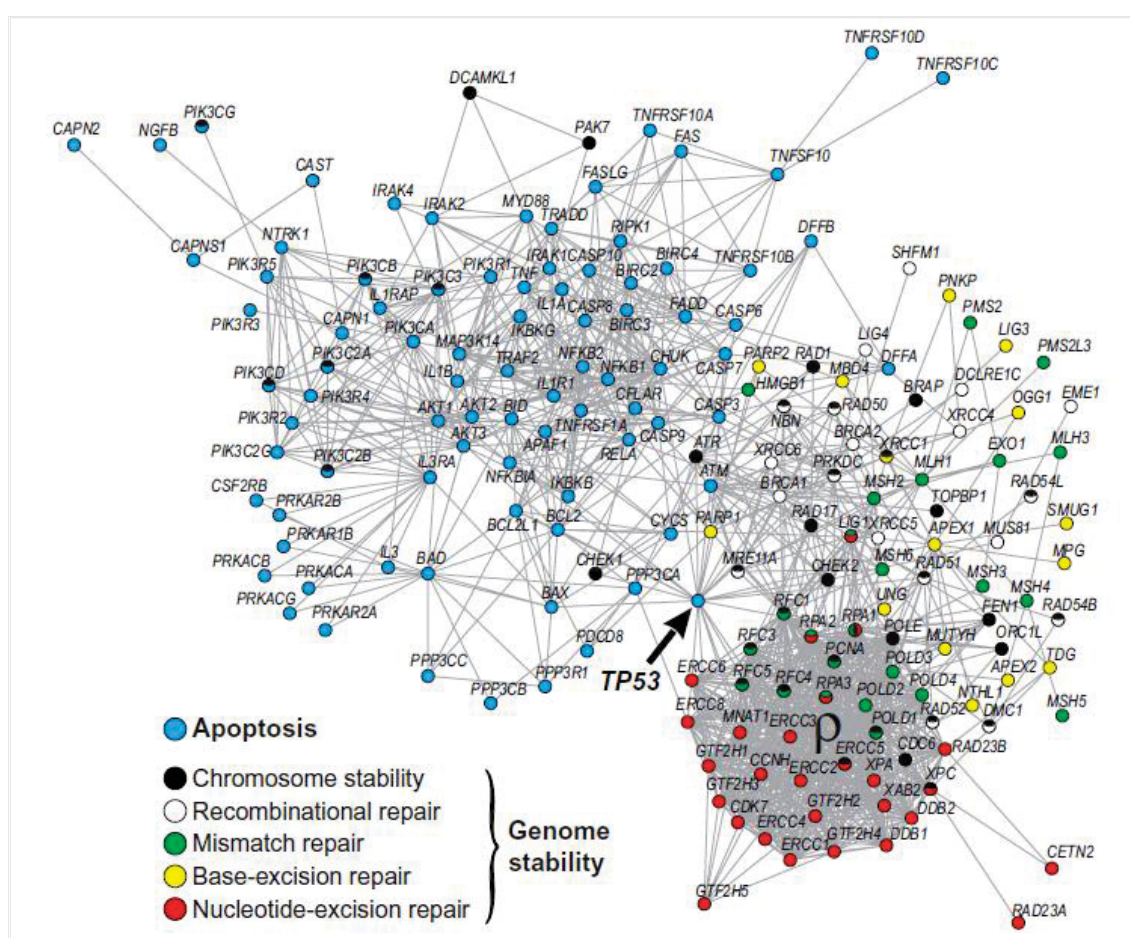
O conceito de apoptose está associada à manutenção da homeostase dos tecidos (Danial; Korsmeyer, 2004). A morte celular programada (Programmed Cell Death - PCD) na perspectiva de organismos multicelulares garante a substituição de células velhas e disfuncionais, que estão comprometidas pela acumulo de danos celulares devido a fatores ambientais, bem como participa diretamente no desenvolvimento do tecido (Lettre; Hengartner, 2006). De acordo com o KEGG (Kyoto Encyclopedia of Genes and Genomes - KANEHISA et al., 2006), base de dados de referência para vias, até 100 genes trabalham coordenadamente em apoptose. A remoção de um destes componentes afeta vários outros e podem prejudicar toda a via. Em organismos metazoários complexos, a apoptose defeituosa está associada com desordens da organogênese e crescimento celular descontrolado, comportamento característico em doenças neoplásicas (Hipfner; Cohen, 2004). Na perspectiva de uma célula de cancro, a apoptose suprimida é uma exigência, a fim de melhorar a aptidão e sucesso da célula (Crespi; Summers, 2005). De certa forma, a apoptose está relacionada com a instabilidade do genoma, no sentido de que há mutações propensas em clones que contém conteúdo genético aberrante, ou seja, um alto número de aberrações cromossômicas e mutações pontuais no DNA precisam de uma apoptose deficiente para escapar da morte celular (Yan et al., 2006; Zhivotovsky; Kromer, 2004).

Mecanismos de manutenção e estabilidade do genoma estão intimamente ligados à componentes apoptóticos, como indica o número elevado de proteínas que interagem com a proteína supressora de tumor TP53. Na verdade, esta proteína interage com os quatro principais mecanismos de reparo do DNA: reparo por excisão de nucleotídeo (Nucleotide Excision Repair - NER), reparo por excisão de bases (Base Excision Repair - BER), reparo de pareamentos errados (Mismatch Repair - MMR), reparação de recombinação (Recombinational Repair - RER) e também com o reparo de recombinação de homólogos (Homologous Recombinational Repair - HRR) e junção de extremidades não homóloga (Non-Homologous End Joining - NHEJ). No que diz respeito a NER e MMR, o gene *TP53* pode atuar em ambas as formas de transativação, dependente e independente (Sengupta; Harris, 2005). Além disso, várias proteínas de reparo do DNA podem



estimular a apoptose em resposta a lesões, como a poli-proteína associada BER (ADPribose), polimerase-1 (*PARP1*) (Alano; Ying; Swanson, 2004) e as proteínas de *Mismatch Repair*, *MSH2*, *MLH1* e *MSH6* (Duckett et al., 1999). De fato, a sobreposição entre a apoptose e genes de reparo do DNA torna difícil uma definição precisa dos limites funcionais entre todos esses sistemas, o que é uma característica de redes biológicas complexas (Barabasi; Oltvai, 2004).

As interações entre os mecanismos de apoptose e de manutenção do genoma têm sido extensivamente documentados e incluem funções de transativação independentes e dependentes, em que a proteína p53 funciona como um 'nó' molecular em relação ao mecanismo de resposta ao dano do DNA (**Figura 9**), essencial para organismos multicelulares complexos, sendo que falhas neste sistemas estão diretamente associadas com a instabilidade do genoma observada em câncer (Castro et al., 2008).



**Figura 9.** Rede de apoptose e processos envolvidos na estabilidade do genoma presente no estudo de Castro et al., 2008 (adaptado de Castro et al., 2008).

A abordagem de construção de uma rede no contexto da estabilidade do genoma visa analisar o funcionamento integrado dos genes em relação à um panorama de interesse, com isso é possível sobrepor camadas de informações distintas e obter indícios conceituais do comportamento biológico em amostras de interesse.



### **3. OBJETIVOS**

#### **3.1 OBJETIVOS GERAIS**

O presente trabalho teve como objetivo a construção de um grafo de rede conceitual dos genes envolvidos na estabilidade do genoma disposto em vias representativas, utilizando-a para avaliar a expressão diferencial em amostras de câncer de colorretal e amostras relacionadas com o seu desenvolvimento.

#### **3.2 OBJETIVOS ESPECÍFICOS**

Os objetivos específicos incluem as seguintes etapas:

- Recuperar interações dos genes da rede com base em suas relações proteicas;
- Verificar a atividade da rede, por análise de expressão diferencial, em amostras de câncer colorretal em relação ao tecido normal, ou com um perfil distinto do tecido de interesse;
- Verificar a atividade dos sistemas de reparo, apoptose, ciclo celular e inflamação na rede de genes de estabilidade do genoma;
- Utilizar um método estatístico para confirmar a análise de atividade da expressão diferencial.

## 4. MATERIAIS E MÉTODOS

### 4.1 OBTENÇÃO DE DADOS PARA CONSTRUÇÃO DA REDE

Os dados foram coletados na base de dados Ontocancro 2.0, a qual contém vias de genes envolvidos nos mecanismos de manutenção do genoma, compreendendo, mecanismos de reparo do DNA, do ciclo celular, apoptose e senescência. Devido à grande heterogeneidade dos dados disponíveis em todas as demais bases, como por exemplo as plataformas do NCBI (National Center for Biotechnology Information) e Gene Ontology, e a ausência de uma ferramenta especializada nas vias do mecanismo de reparo do DNA, a Ontocancro visa facilitar a modelagem e integração destes dados.

Os dados de pathways do Ontocancro 2.0 foram anotados em uma tabela de 8 colunas por 1105 linhas. A primeira coluna está anotada para nomenclatura de Approved Symbol, nas demais sete colunas foram determinadas as informações de vias (pathways), elas são, Apoptosis, Inflammation, DNA Damage, Cell Cycle, Apoptosis/Inflammation, DNA Damage/Cell Cycle, Intersection. A primeira linha foi determinada como cabeçalho da tabela, e as demais contém a informação dos genes presentes nos mecanismos de estabilidade do genoma (ANEXO1).

O processo de atribuição de função aos genes foi realizado pela sua presença nas vias (pathways), onde os que estavam presentes em uma das quatro principais vias ou nos subgrupos de intersecção, eram anotados de acordo com a combinação entre sua presença e ausência. Foram marcados como presente (1), e ausentes (0). Os genes que se apresentavam como únicos em uma via, foram marcados exclusivos, os que possuíam dupla função em processos paralelos como por exemplo apoptose e inflamação, e dano ao DNA e ciclo celular, foram enquadrados em subgrupos menores, descritos acima. Os demais que não se enquadraram em nenhum dos grupos principais e não possuíam ligações paralelas, e compartilhavam ligações com outros módulos da rede ou interação com mais de dois módulos foram chamados de genes de intersecção (Intersection).

A utilização dessa abordagem permite recuperar informações sobre as funções desempenhadas por um determinado gene e dividi-la em diferentes processos de modo que sirvam para a recuperação da interação entre genes, assim

os dividindo em grupos. Essa divisão foi utilizada para a recuperação das interações gênicas, e servirá para medir quão próximos ou distantes estão, os processos celulares entre si e a interação entre os genes individualmente.

#### 4.2 RECUPERAÇÃO DA INTERAÇÃO ENTRE PROTEÍNAS

Com a obtenção da nomenclatura dos genes e a categorização dos processos aos quais pertencem, pode-se estabelecer suas interações. A obtenção desses dados é de extrema importância para a disposição das vias, que conceitualmente correspondem aos processos celulares envolvidos na estabilidade celular à partir da estabilidade do genoma.

Para a estabelecer as interações entre os genes presentes na base de dados Ontocancro 2.0, foi usado o banco de dados STRING v. 9.1. O STRING é um banco de dados de interações proteicas conhecidas e previsíveis. Suas interações são baseadas em associações diretas (físicas) e indiretas (funcionais), derivadas à partir de quatro fontes: *Genomic Context*, *High-throughput Experiments*, *Coexpression (Conserved)*, *Previous Knowledge (Database)*. O STRING integra quantitativamente dados de interação dessas fontes. No trabalho foram utilizados dados de interação proteína/proteína somente em *Homo sapiens* e todo o processo de recuperação das interações e as demais fases deste trabalho foram feitas na linguagem de programação e ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos, R, especificamente em seu ambiente de desenvolvimento integrado, R-Studio, utilizando o pacote do STRING para ambiente R, “*STRINGdb*” (Franceschini et al., 2013) presente no repositório Bioconductor, que abriga softwares de desenvolvimento aberto para análise e compreensão de dados genômicos gerados por experimentos em biologia molecular. As evidências do STRING escolhidas para este trabalho foram, *High-throughput Experiments*, *Database (Previous Knowledge)* e *Coexpression*. Para combinar essas evidências foi implementada uma função de combinação de *scores*, gerando um nível de probabilidade de confiança alto de 0,9, ou seja, baseado na combinação das evidências obtivemos noventa por cento de chances em que dois genes que codificam uma proteína funcional estão ligados por interações em humanos, e baseado nesse *score* podemos quantificar quão próximas são as ligações entre dois genes. Como objeto de saída obteve-se uma matriz com

a anotação de genes e suas respectivas vias, onde todas as suas interações foram mapeadas.

#### 4.3 CONSTRUÇÃO DA REDE

Para construir um grafo de interação e formar uma rede conceitual foram utilizados os dados de nomenclatura e função dos genes, juntamente com suas interações. Este grafo consiste em unir informações em um elemento visual, onde os nós representam os genes e vias. O nó possui a informação do gene e por cores pode-se dividi-los em vias, as quais representam os processos de interesse que serão analisados. Os links ou caminhos representam as interações recuperadas das proteínas codificadas pelos genes selecionados. Com isso temos um elemento gráfico que poderá ser utilizado como camada inicial para análise funcional de expressão gênica.

Para a confecção e remodelagem da rede de estabilidade do genoma foi utilizado o pacote depositado no Bioconductor, “*RedeR*” (Castro MAA et al., 2012). O “*RedeR*” é um pacote em R combinado com a linguagem JAVA para a representação de redes modulares. Sua função é representar estruturas modulares, redes, e múltiplos níveis hierárquicos de associações.

#### 4.4 OBTENÇÃO E ANÁLISE DOS DADOS DE EXPRESSÃO GÊNICA

O conjunto de microarranjos (*microarrays*) de expressão gênica dos diferentes tipos de fenótipos de tecido de colorretal foi obtido do repositório de dados de expressão gênica GEO “*Genome Expression Omnibus*”. O conjunto escolhido está depositado na plataforma “GPL570”, plataforma que contém o array de anotação do genoma humano da Affymetrix ([HG-U133\_Plus\_2] Affymetrix Human Genome U133 Plus 2.0 Array). As séries selecionadas foram a “GSE4183” e “GSE4554”. A série “GSE4183” possui amostras de biopsias de colo com o fenótipo normal, e com os fenótipos de adenoma, câncer de colorretal (CRC), e doença do intestino inflamado (IBD) em um grupo de 53 amostras (Galamb et al., 2008). A série “GSE4554” possui amostras de tumores de colorretal retiradas de 84 pacientes, alguns com o fenótipo

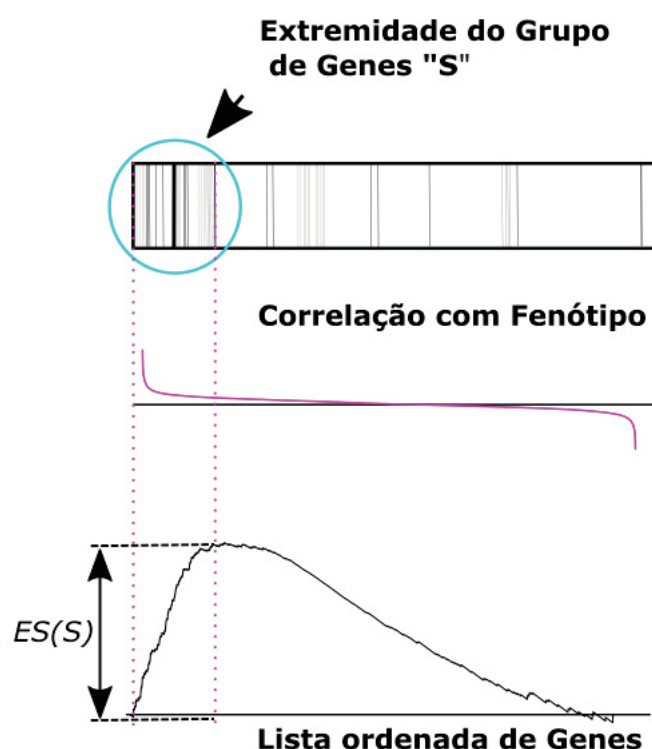
de estabilidade microssatélite (MSS) e outros com o perfil de instabilidade microssatélite (MSI) em um grupo de 84 amostras (Watanabe et al., 2006).

A análise de expressão diferencial consiste na identificação dos genes diferencialmente expressos entre os diferentes fenótipos. Com o auxílio do pacote “*limma*” (Smyth, 2005) para o ambiente estatístico R. Limma é um pacote para a análise de dados de expressão gênica provenientes de microarray ou RNA-Seq. Uma capacidade central é a utilização de modelos lineares para avaliar a expressão no contexto de experimentos multifatoriais. Possui características que tornam as análises estáveis mesmo para experimentos com pequeno número de matrizes, e isso é obtido através das informações dos genes. Isto é especialmente concebido para analisar experiências complexas com uma variedade de condições experimentais de preditores. O modelo linear e as funções de expressão diferencial são aplicáveis a qualquer dados quantitativos, incluindo microarrays, RNA-seq e PCR quantitativa.

Utilizando esta ferramenta, foram elencados os genes diferencialmente expressos relativos a cada fenótipo (adenoma, CRC, IBD, MSI), para um valor de  $p=0.01$ , utilizando como método de ajuste o *Adjusted P-value*.

#### 4.5 ANÁLISE DE ENRIQUECIMENTO DE GRUPO DE GENES

O GSEA - sigla em inglês para análise de enriquecimento de grupos de genes - *Gene Set Enrichment Analysis* (Subramanian et al., 2005), verifica dentre um subgrupo dentro de uma agrupamento de genes (GS's) fornecidos *a priori*, os quais mantêm um padrão de distribuição (**Figura 10**) de genes estatisticamente relacionados com dois estados biológicos. Neste caso amostras de tecido de colo normal vs. Adenoma, CRC, IBD e o fenótipo de MSS vs. MSI.



**Figura 10.** Ilustração representativa do método de GSEA: Os dados de expressão são ordenados pelo seu grau de correlação com o grupo biológico analisado. Então é computada um somatório que percorre a lista de genes e a cada vez que a posição analisada corresponde a um gene do grupo incrementa-se o valor de enriquecimento nesse gene (S) e decrementa caso o efeito seja ao contrário. O índice de enriquecimento (ES) de um grupo de genes “S” é dado pelo maior valor atingido por esta soma. FONTE: Adaptado de Subramanian et al., 2005.

A análise de GSEA foi feita no ambiente R com o auxílio do pacote “*HTSanalyzer*” (Wang et al., 2011), que enquadra os grupos de genes, realiza e retorna os dados provenientes da análise estatística, além de gerar os gráficos representativo da GSEA. Com essa análise foi possível comparar os resultados de expressão diferencial que foram gerados pelo processo de visualização gráfica

#### 4.6 VISUALIZAÇÃO GRÁFICA DA ATIVIDADE FUNCIONAL DA REDE DE ESTABILIDADE DO GENOMA

A visualização dos dados foi feita utilizando o software viaComplex© (Castro et al., 2008-2009). O viaComplex© é um software de análise topográfica e estabelece o panorama transcricional de redes, utilizando-se de valores médios de

cada um dos grupos vs. a média dos controles. Esta ferramenta permite utilizar uma rede de genes como pano de fundo para avaliar o “espalhamento” transcricional, ou seja, o efeito da alteração transcricional sobre a rede, aplicando uma terceira dimensão de cor sobre as duas dimensões da rede. Assim obtivemos uma figura topográfica com as partes da rede que estão alteradas com relação ao controle de cada amostra, tanto enriquecidas quando diminuídas em relação ao fenótipo de controle.

## 5. RESULTADOS

### 5.1 REDE DE ESTABILIDADE DO GENOMA

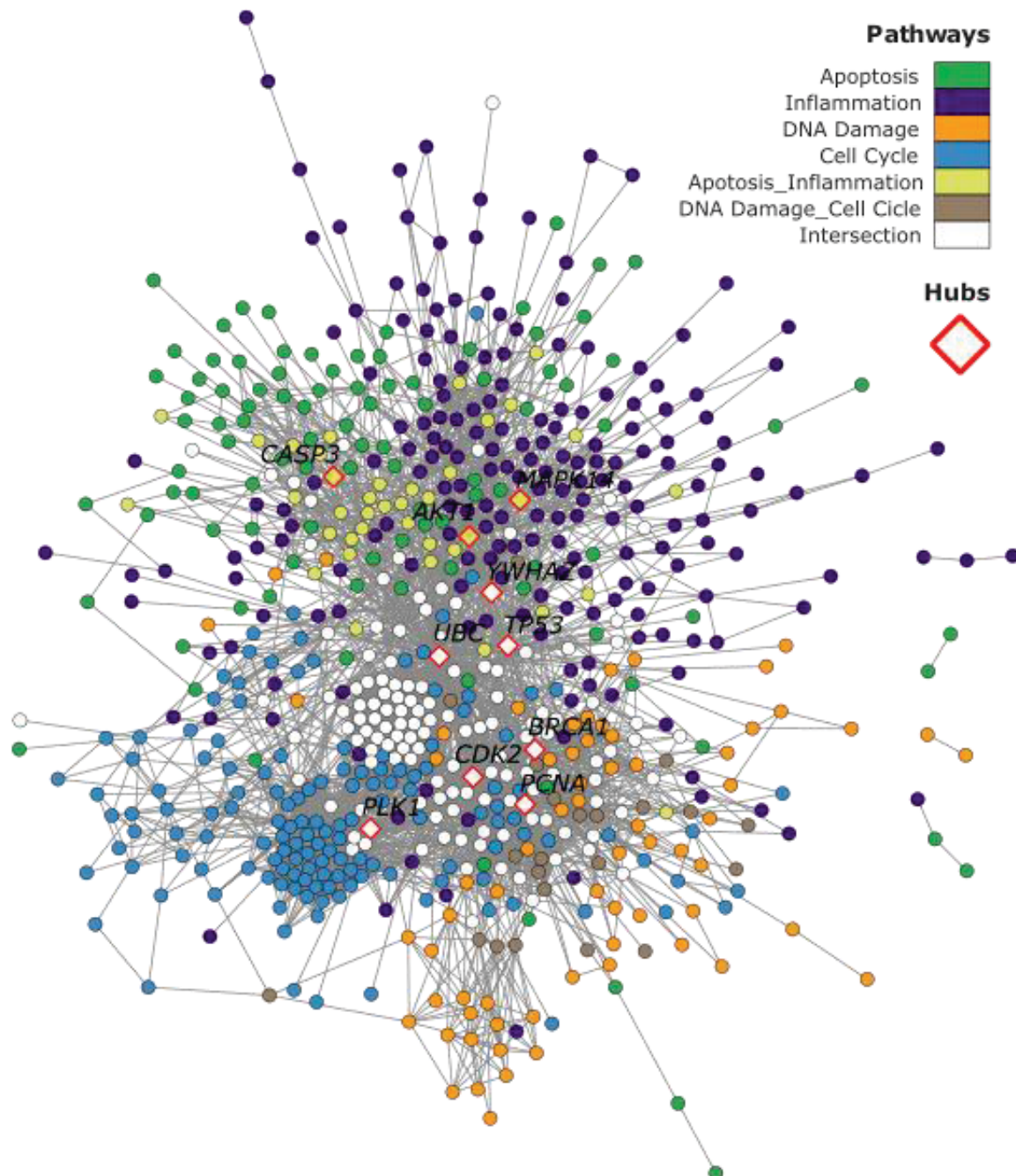
A rede de estabilidade do genoma (**Figura 11**) serviu como plano de fundo para as análises funcionais do estudo. Como descrito anteriormente os atributos gráficos da rede são: nós ou nodos e caminhos ou vértices (links). Cada nó da rede corresponde a um gene codificador de uma proteína funcional, e os links correspondem às interações que os nodos (genes) possuem entre si, ou seja, quanto maior for o a interação entre duas proteínas, a tendência é que dois nós se agrupem topologicamente na rede, e quanto menor a interação tendem a se afastar.

A disposição da rede nesta forma gráfica é dada pelo conceito de mola e repulsão utilizada pelo pacote “*RedeR*” (Castro MAA et al., 2012), onde cada borda pode ser considerado como uma mola com um comprimento. Um determinado alvo, pode exercer uma força de repulsão ou de atração sobre os nós conectados, enquanto que os nós são análogos às cargas mutuamente repulsivas que se movem de acordo com a força aplicada neles.

Baseado nesses conceitos descritos acima, foi possível agrupar os genes de acordo com suas funções nas vias (pathways) ao qual eles correspondem. As vias formam módulos que ilustram quatro principais processos celulares, são eles: apoptose (Apoptosis), inflamação (Inflammation), genes de estabilidade e reparo ao dano do DNA (DNA Damage) e ciclo celular (Cell Cycle). Os genes que possuem função dupla em dois processos que estão mais próximos topologicamente foram destacados como: apoptose/inflamação (apoptosis\_inflamação), dano ao DNA/ciclo celular (DNA Damage\_Cell Cycle), e aqueles que desempenham funções entre dois módulos topologicamente distintos ou função em pelo menos três vias da rede foram destacados como genes de intersecção (Interscetion). Também foram destacados com um *frame* vermelho em formado de balão 10 genes considerados *hubs* nessa rede, e para obtê-los foi utilizado a medida de centralização *betweenness* (intermediação). A intermediação é uma medida de centralidade de um nó em uma rede. Ela é igual ao número de menores caminhos de todos os vértices para quaisquer outros vértices que passam por aquele nó, em outras palavras são os nós que mais conectam a rede.



## GRAFO DA REDE DE ESTABILIDADE DO GENOMA



**Figura 11.** Rede de estabilidade do genoma ampliada com os genes (nós) presentes na base de dados do Ontocancro 2.0 e sua anotação para um determinado processo biológico celular (via). Cada nó representa um gene que codifica uma proteína funcional e os processos biológicos estão destacados por uma cor. Apoptose (verde), Inflamação (roxo), Reparo do DNA (laranja) e Ciclo Celular (azul). Genes com dupla função em Apoptose e Inflamação (verde musgo), e em Reparo do DNA e Ciclo Celular (marrom). Genes de Intersecção, com mais de duas funções em grupos distintos dos citados acima estão destacados em branco. Hubs da rede circundados em balões vermelhos. As interações entre os nós foram recuperadas pelas evidências presentes no banco de dados STRING, baseadas em interações proteína/proteína.

O conceito de *Hub* é descrito nas ciências de redes como um nó que possui um número de links que excede a média dos demais nós, ou seja, em uma rede o *hub* funciona como peça central, que recebe os sinais transmitidos pelas vias ou módulos e os retransmite para os demais nós (Barabási, 2014).

Os *hubs* encontrados na rede foram *TP53*, *UBC*, *BRCA1*, *CDK2*, *PCNA*, *PLK1*, *YWHAZ*, *AKT1*, *MAPK14*, *CASP3*.

O gene *TP53*, responsável pela produção da proteína supressora de tumor *P53* é destacado pelo estudo realizado por Castro *et. al.* (2008). Segundo Castro *et. al.* (2008) *TP53* funciona como um ‘nó’ molecular em relação ao mecanismo de resposta ao dano do DNA, e está diretamente ligada com funções de transativação independentes e dependentes na rede de estabilidade do genoma.

A proteína produzida pelo gene *UBC* - Ubiquitina C tem importante papel na degradação de proteínas, pois realiza marcações para que sejam degradadas, e está ligada diretamente a fases do ciclo celular, como também possui importante papel nos mecanismos de reparo e de apoptose.

*BRCA1* - *Breast Cancer 1*, coordena uma gama diversificada de vias celulares, como reparo ao dano do DNA, ubiquitinação e regulação da transcrição para manter a estabilidade genômica. Mede a atividade ubiquitina ligase necessária para sua função supressora de tumores. Desempenha um papel central no reparo do DNA, facilitando a resposta celular.

O gene Quinase dependente de ciclina 2 - *CDK2* está envolvido no controle do ciclo celular.

*PCNA* – Antígeno nuclear de proliferação celular, é auxiliar a DNA polimerase delta e está envolvido no controle da replicação e no processo de *mismatch repair*.

*PLK1*, atua na inativação dos inibidores do gene *APC*, um dos primeiros genes a sofrer mutação na diferenciação celular carcinogênica.

*YWHAZ*, está envolvido nos sinais de stress celular e está envolvido nos processos de reparo do DNA.

O gene que codifica a enzima *AKT1* induzida por fatores de crescimento, age como oncogene.

*MAPK14* atua em processos de stress e inflamação e *CASP3* – Caspase 3 atua na cascata de caspase, na via intrínseca da apoptose, nos sinais de receptores

para morte programada e está diretamente ligada aos demais processos que englobam a apoptose.

Todos os *hubs* demonstrados estão descritos na literatura e possuem importante papel nas funções biológicas dos genes presentes na rede de estabilidade do genoma.

## 5.2 ANÁLISE DE EXPRESSÃO GLOBAL

A análise de expressão global foi padronizada da seguinte forma, primeiramente a rede de estabilidade do genoma foi redimensionada para coordenadas válidas no Viacomplex©, e como dado de saída foi gerada uma imagem de expressão diferencial da rede, com os valores de expressão de cada gene que somados formam o padrão de expressão das vias. A expressão diferencial é resultado das amostras (adenoma, CRC, IBD, MSI) em relação à um controle. Após obter a imagem baseada nos valores de expressão das vias, foram utilizados os mesmos dados para o teste de GSEA, onde foi possível observar o mesmo padrão de expressão diferencial obtido na figura gerada pelo Viacomplex© em todas as amostras. Para demonstrar os resultados, foi formado um fluxo de análise que consiste; em uma figura da rede redimensionada, uma figura da expressão diferencial das vias da rede em cada uma das amostras utilizadas, uma tabela com os dados obtidos na GSEA em um teste de valor-*p* e dois gráficos, um contendo um ranking de genes e sua alteração em relação à um controle, e outro com o score de enriquecimento. (**Figuras 12,13,14,15**).

### 5.2.1 DOENÇA INFLAMATÓRIA INTESTINAL (IBD)

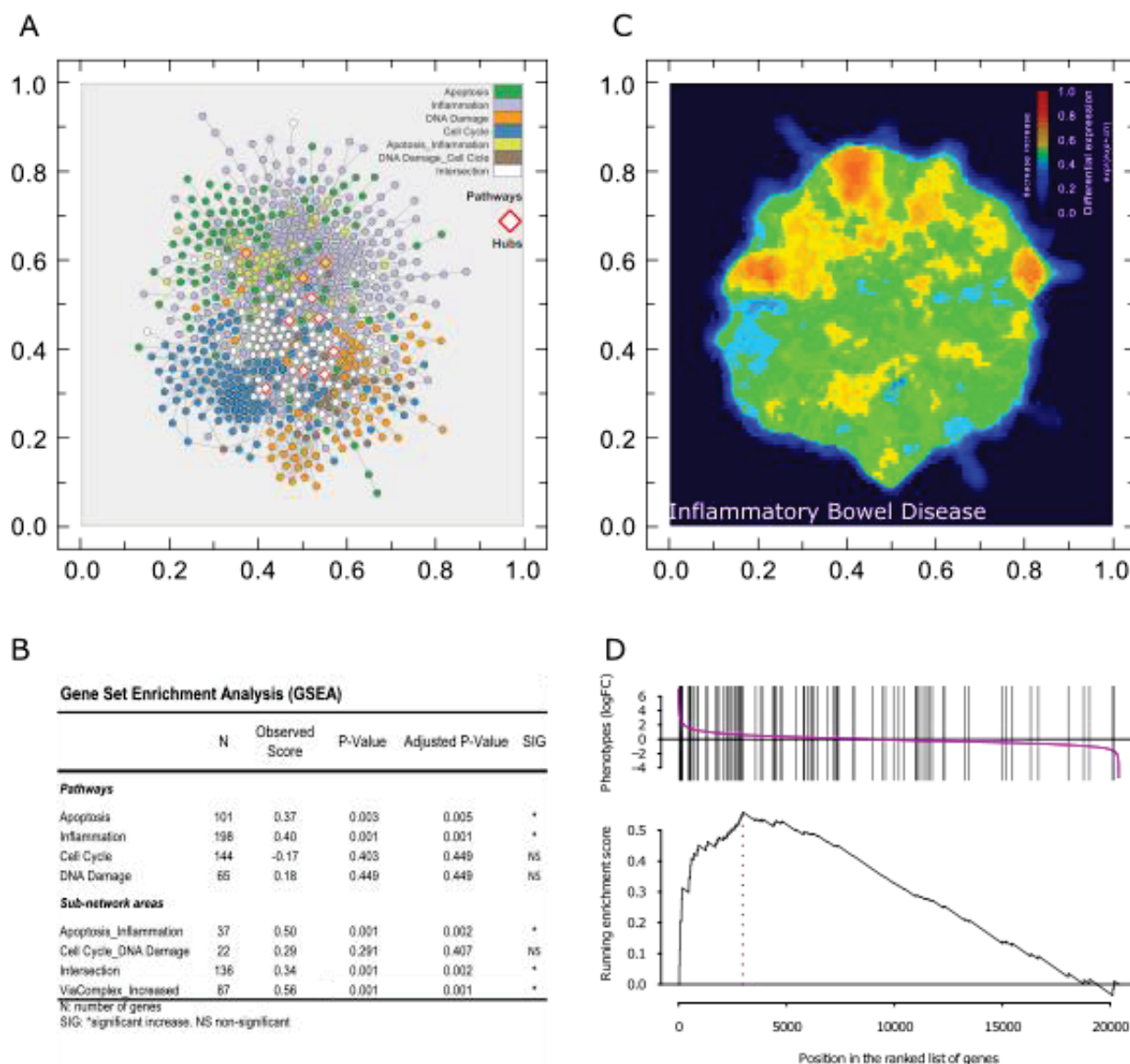
Na **Figura 12**, a análise visual topográfica realizada pelo software Viacomplex©, sugere uma expressão diferencial aumentada nos módulos de inflamação e apoptose da rede de estabilidade do genoma, no tecido afetado pela doença com relação ao tecido normal de colo (controle).

A estatística fornecida pelo teste de GSEA na amostra do fenótipo da doença do intestino inflamado vs. tecido normal de colo mostrou-se significativa nos módulos de inflamação e apoptose, com o significância do valor ajustado de  $p=0.001$  para

inflamação, e valor ajustado de  $p=0.005$  em apoptose. O subgrupo de genes que desempenham função dupla tanto em apoptose quanto na inflamação também se mostraram significativos com um valor ajustado de  $p=0.002$ . Os genes de intersecção também foram significativos para essa amostra com o valor ajustado de  $p=0.002$ .

Para a confecção do gráfico de GSEA foram selecionados somente os genes dos módulos que se mostraram significativamente expressos na análise estatística e pelo Viacomplex©. O gráfico de *score* de enriquecimento mostrou que existe um grupo de genes com o fenótipo alterado na amostra da doença do intestino inflamado com relação ao tecido normal de colo, demonstrando a veracidade dos resultados comentados acima.

## FLUXO DE ANÁLISE FUNCIONAL, ESTATÍSTICA E GRÁFICA EM IBD



**Figura 12.** Fluxo de análise gráfica funcional em IBD (GSE4183). (A) Rede de genes envolvidos nos processos de estabilidade do genoma humano redimensionada para análise topográfica de acordo com as coordenadas do software ©Viacomplex. Em verde está representado o módulo de apoptose; em magenta, inflamação; laranja, dano ao DNA; azul, ciclo celular; verde musgo, genes funcionais tanto em apoptose como em inflamação; marrom, genes funcionais no dano ao DNA e ciclo celular; branco, genes funcionais em outras combinações de vias, presentes em dois ou mais nos processos. (B) Tabela com os dados da análise de enriquecimento de grupo de genes (GSEA) das vias descritas e suas subáreas. (C) Imagem da atividade de expressão diferencial em IBD vs. tecido normal(controle). (D) Gráficos do método de GSEA, demonstrando tanto expressão diferencial a partir de um ranking de genes alterados em IBD quanto o score de enriquecimento da distribuição do grupo de genes diferencialmente expressos em C obtidos da análise prévia no Viacomplex.



O padrão demonstrado pela amostra de IBD nos deu indício de que nosso fluxo de análise é coerente, pois esperava-se que uma amostra de doença inflamatória estivesse alterada em relação a um tecido normal.

A lesão ocasionada por IBD é uma lesão benigna, por isso a atividade dos genes de reparo do DNA e do ciclo celular ainda não estão diferencialmente expressos em relação a um tecido normal, isto sugere que as funções de reparo e ciclo celular estão funcionando neste tipo de doença.

Nota-se uma variação significativa em apoptose, provavelmente pelo fato do tecido de colo sofrer um aumento na inflamação, onde existe um grau de substituição e degradação celular elevada.

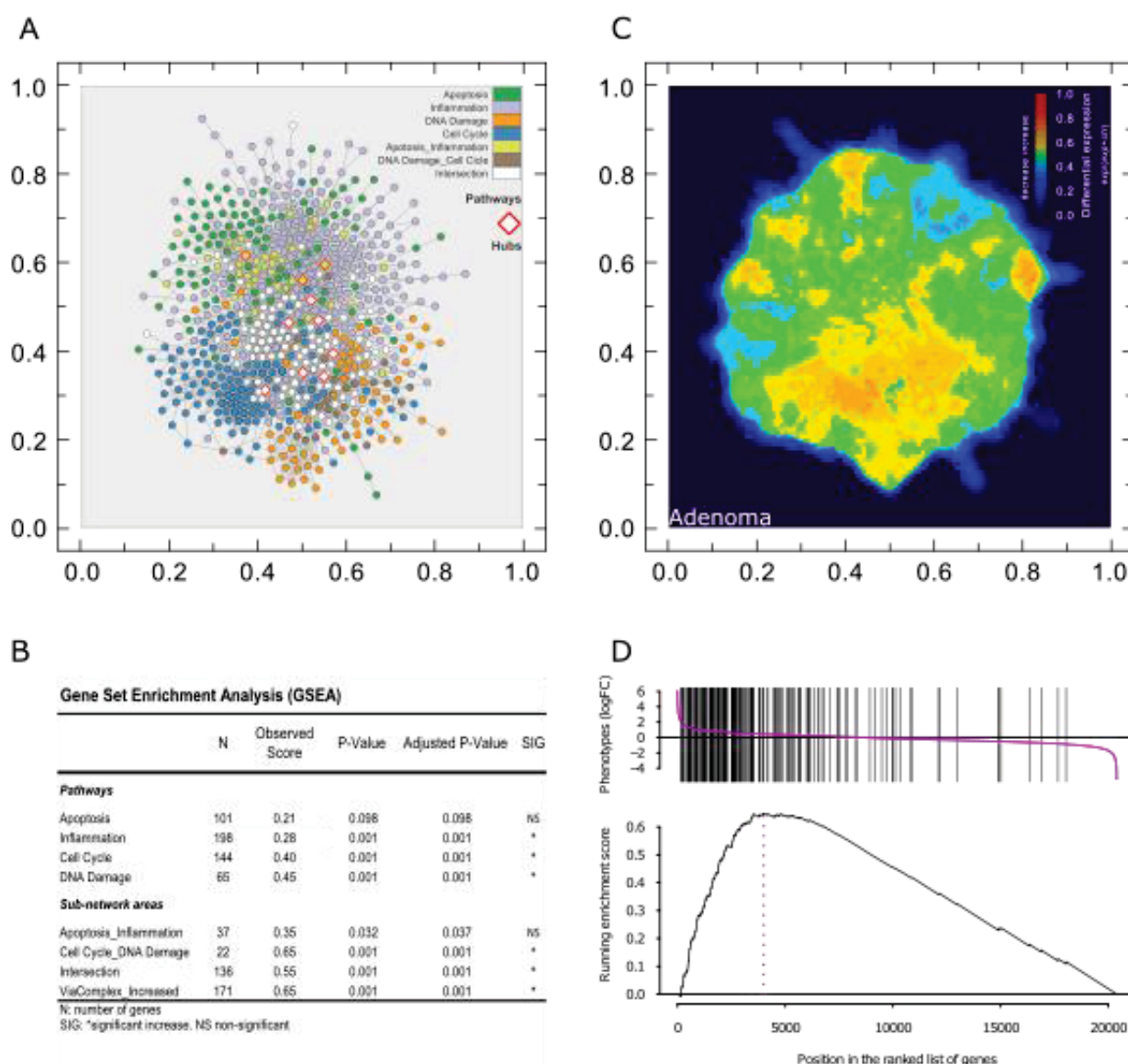
### 5.2.2 ADENOMA

Em adenoma (**Figura 13**), a análise topográfica do Viacomplex© demonstra um panorama diferente de IBD, mesmo a inflamação estando diferencialmente expressa. Genes de ciclo celular e de reparo ao dano do DNA também estão alterados em adenoma. Nas subáreas da rede, genes de dupla função em ciclo celular e reparo de dano ao DNA, juntamente com os genes de intersecção sugerem uma expressão significativa.

O teste estatístico de GSEA reforça a assinatura de expressão observada no Viacomplex©, em que a inflamação, ciclo celular, e reparo ao dano do DNA mostram significância de valor ajustado de  $p=0.001$ , assim como os genes do modulo de dupla função em ciclo celular e reparo ao dano do DNA e os genes de intersecção. O gráfico de GSEA nos mostra que o grupo de genes aumentado na análise topográfica também está alterada com relação ao fenótipo normal, onde este agrupamento demonstra o mesmo score de enriquecimento dos genes de ciclo celular/dano ao DNA que é igual a 0.65.

Adenoma é uma lesão pré-maligna, onde existe acúmulo de células cerosas no epitélio da mucosa do colo, e isso sugere o aumento nos sistemas de reparo e ciclo celular, e esse aumento pode ser funcional, em respostas aos danos, ou pode ser consequência de danos específicos nesse grupo de genes relacionados com os processos de ciclo celular e de reparo ao dano do DNA.

# FLUXO DE ANÁLISE FUNCIONAL, ESTATÍSTICA E GRÁFICA EM ADENOMA



**Figura 13.** Fluxo de análise gráfica funcional em adenoma (GSE4183). (A) Rede de genes envolvidos nos processos de estabilidade do genoma humano redimensionada para análise topográfica de acordo com as coordenadas do software ©Viacomplex. Em verde está representado o módulo de apoptose; em magenta, inflamação; laranja, dano ao DNA; azul, ciclo celular; verde musgo, genes funcionais tanto em apoptose como em inflamação; marrom, genes funcionais no dano ao DNA e ciclo celular; branco, genes funcionais em outras combinações de vias, presentes em dois ou mais processos. (B) Tabela com os dados da análise de enriquecimento de grupo de genes (GSEA) das vias descritas e suas subáreas. (C) Imagem da atividade de expressão diferencial em adenoma vs. tecido normal (controle). (D) Gráficos do método de GSEA, demonstrando tanto expressão diferencial a partir de um ranking de genes alterados em adenoma quanto o score de enriquecimento da distribuição do grupo de genes diferencialmente expressos em C obtidos da análise prévia no Viacomplex.

### 5.2.3 CANCER DE COLORRETAL (CRC)

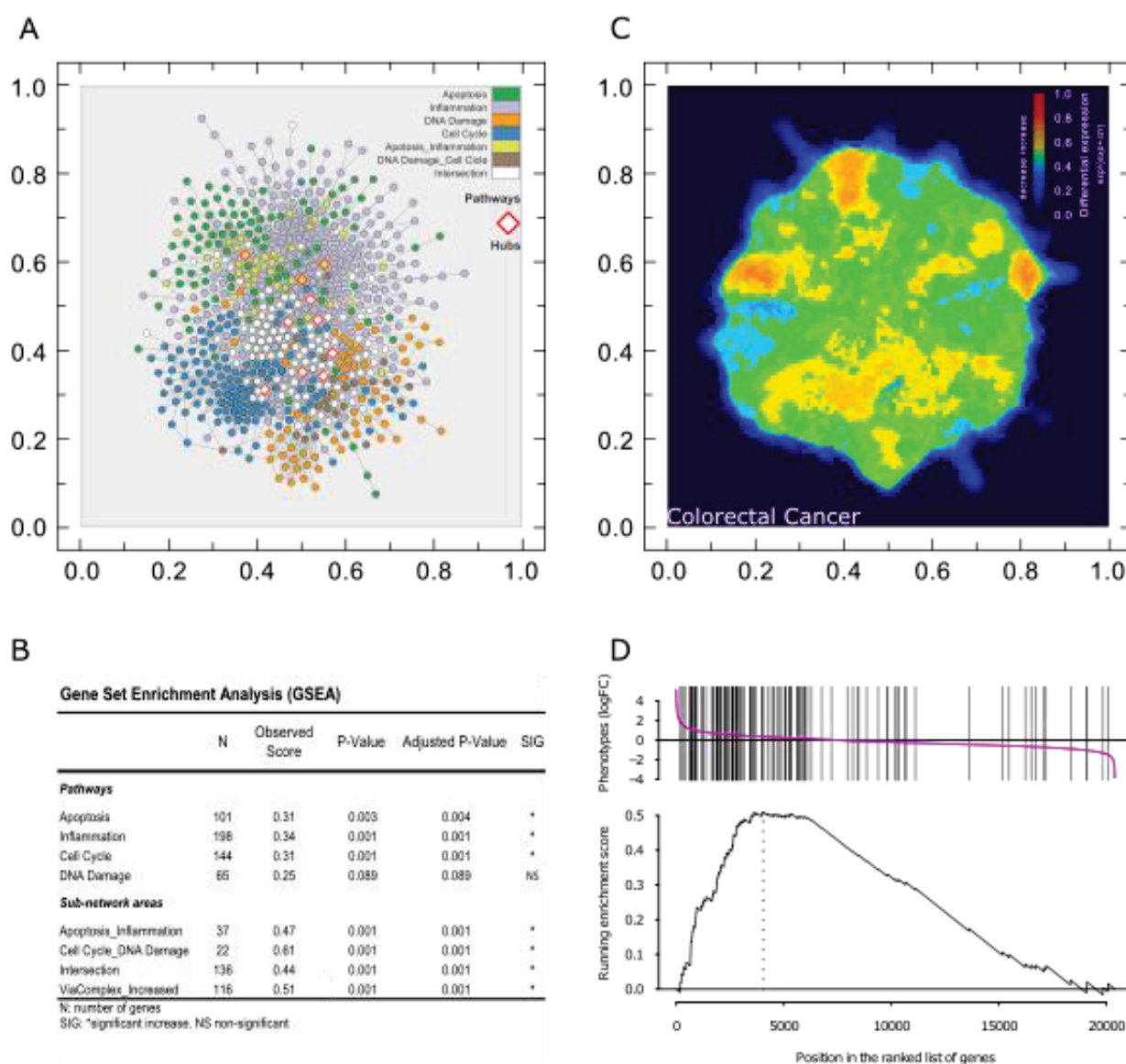
Na amostra de câncer de colorretal (**Figura 14**) o panorama topográfico mostrou aumento de expressão nos genes em apoptose, inflamação e ciclo celular, sendo que há decréscimo na expressão dos sistemas de reparo se comparado a uma lesão pré-maligna de adenoma.

A estatística da GSEA mostra significância ajustada de  $p=0.004$  em apoptose,  $p=0.001$  em inflamação e ciclo celular, enquanto os genes de reparo ao dano do DNA se mostraram não significativos. Os genes de função dupla tanto em apoptose/inflamação e ciclo celular/reparo ao dano do DNA se mostraram significativos, assim como os genes de intersecção, com o valor ajustado de  $p=0.001$ . O gráfico também mostra que há um grupo de genes enriquecido no fenótipo de câncer de colorretal em relação a um fenótipo normal.

Mesmo sendo uma lesão maligna, a amostra de CRC mostra uma menor atividade dos sistemas de reparo se comparado à adenoma, apesar do aumento dos danos no DNA causada pela desordem dos sistemas celulares. É importante ressaltar que CRC mostrou alguma atividade em genes de apoptose e inflamação. Isto sugere uma mudança nestes mecanismos no estágio de câncer, e essa mudança não significa uma ativação do processo de apoptose em si, mas pode estar envolvida com vias de sinalização, graças aos erros em reparo, fazendo com que a apoptose não esteja funcionando corretamente.



## FLUXO DE ANÁLISE FUNCIONAL, ESTATÍSTICA E GRÁFICA EM CRC



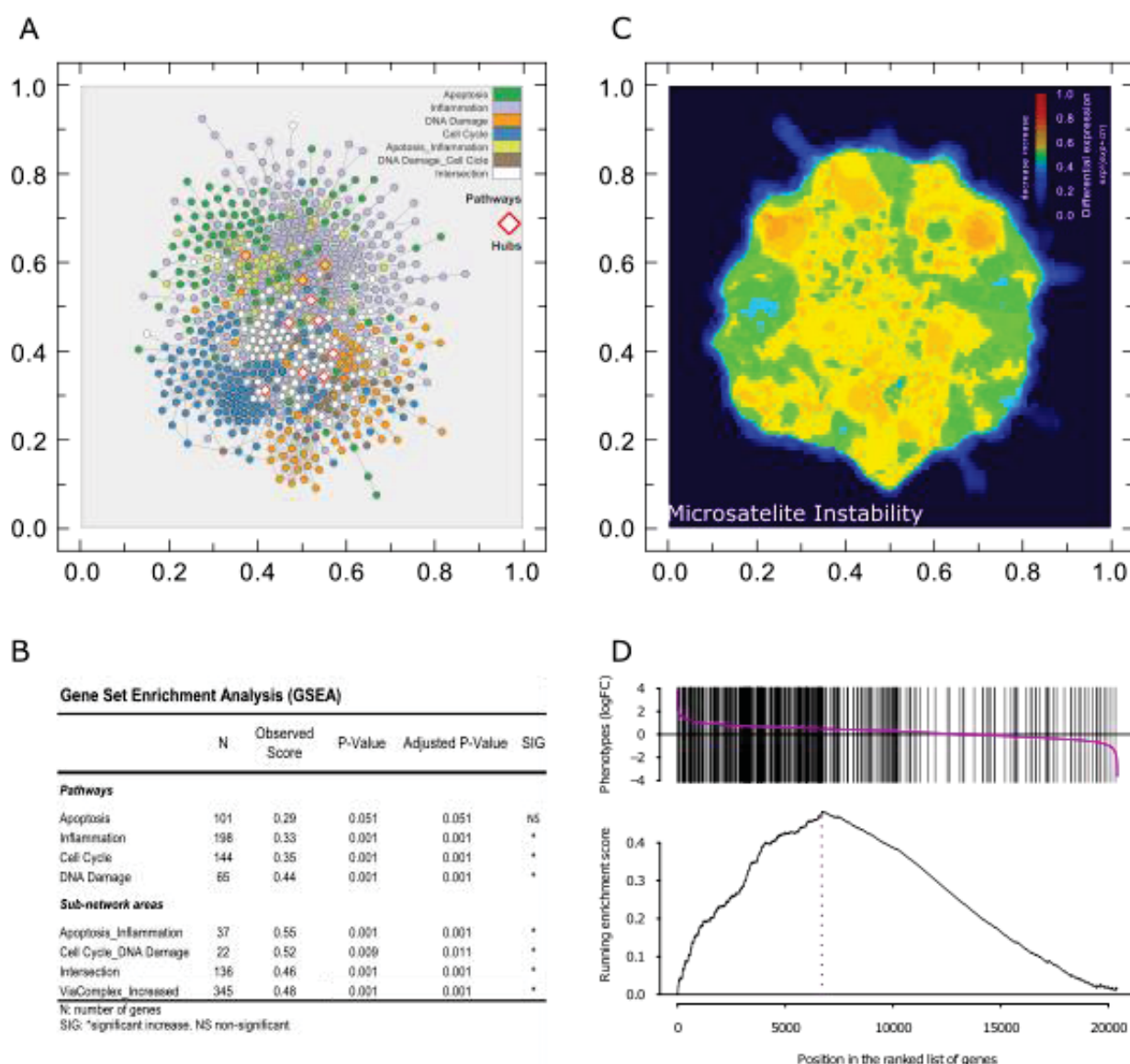
**Figura 14.** Fluxo de análise gráfica funcional em CRC (GSE4183). (A) Rede de genes envolvidos nos processos de estabilidade do genoma humano redimensionada para análise topográfica de acordo com as coordenadas do software ©Viacomplex. Em verde está representado o módulo de apoptose; em magenta, inflamação; laranja, dano ao DNA; azul, ciclo celular; verde musgo, genes funcionais tanto em apoptose como em inflamação; marrom, genes funcionais no dano ao DNA e ciclo celular; branco, genes funcionais em outras combinações de vias, presentes em dois ou mais processos. (B) Tabela com os dados da análise de enriquecimento de grupo de genes (GSEA) das vias descritas e suas subáreas. (C) Imagem da atividade de expressão diferencial em CRC vs. tecido normal (controle). (D) Gráficos do método de GSEA, demonstrando tanto expressão diferencial a partir de um ranking de genes alterados em CRC quanto o score de enriquecimento da distribuição do grupo de genes diferencialmente expressos em C obtidos da análise prévia no Viacomplex.

#### 5.2.4 INSTABILIDADE MICROSSATÉLITE (MSI)

Na instabilidade microssatélite (**Figura 15**) o Viacomplex© mostrou aumento da expressão diferencial em todas as vias da rede, exceto em apoptose.

Na estatística de GSEA todas as vias da rede apresentaram valor ajustado de  $p=0.001$ , exceto apoptose que se mostrou não significativa em MSI. A significância estatística de todas as subáreas da rede foi de  $p=0.001$ , com exceção dos genes do ciclo celular/reparo ao dano do DNA, no valor de  $p=0.009$ . O gráfico de GSEA também mostrou que há um grande número de genes alterados em relação ao fenótipo normal. É importante ressaltar em MSI a quantidade de números de hits presentes no gráfico de GSEA, os quais mostram que uma grande quantidade de genes está alterado nesse fenótipo em relação à células que possuem regiões microssatélites estáveis.

# FLUXO DE ANÁLISE FUNCIONAL, ESTATÍSTICA E GRÁFICA EM MSI



**Figura 15.** Fluxo de análise gráfica funcional em MSI (GSE4554). (A) Rede de genes envolvidos nos processos de estabilidade do genoma humano redimensionada para análise topográfica de acordo com as coordenadas do software ©Viacomplex. Em verde está representado o módulo de apoptose; em magenta, inflamação; laranja, dano ao DNA; azul, ciclo celular; verde musgo, genes funcionais tanto em apoptose como em inflamação; marrom, genes funcionais no dano ao DNA e ciclo celular; branco, genes funcionais em outras combinações de vias, presentes em dois ou mais nos processos. (B) Tabela com os dados da análise de enriquecimento de grupo de genes (GSEA) das vias descritas e suas subáreas. (C) Imagem da atividade de expressão diferencial em MSI vs. tecido normal (controle). (D) Gráficos do método de GSEA, demonstrando tanto expressão diferencial a partir de um ranking de genes alterados em MSI quanto o score de enriquecimento da distribuição do grupo de genes diferencialmente expressos em C obtidos da análise prévia no Viacomplex.

As amostras coletadas da GSE 4554 do GEO são amostras de câncer de colo com padrão de instabilidade microssatélite, e o aumento difuso dos sistemas de reparo e de inflamação juntamente com o aumento significativo das áreas de intersecção entre todos os sistemas da rede, nos sugere um perfil altamente instável quanto a estrutura física e organizacional das funções celulares.

## 6. DISCUSSÃO

### 6.1 CONSIDERAÇÕES GERAIS

O perfil de adenoma, CRC, IBD e MSI ocorre no mesmo tipo de tecido, a mucosa epitelial do intestino, por isso é importante compreender como o desenvolvimento de um tumor maligno nesse tipo de tecido pode estar relacionado nestas diferentes formas de anomalias em colo e colorretal.

O adenoma tem origens em lesões pré-malignas, isso provavelmente explica o padrão observado pelas análises realizadas, em que os sistemas de reparo estão significativamente alterados em relação ao tecido normal de colo, mostrando que existe uma grande correlação dos mecanismos responsáveis pelo reparo ao dano do DNA na diferenciação celular de um tecido normal para um perfil pré-maligno. Esta análise indica que a progressão tumoral está envolvida com mutações e falhas nos processos de reparo e de estabilidade do genoma, e esse panorama sugere o que já era predito na literatura, na medida em que mutações nesses genes alteram o funcionamento do sistema celular, o acúmulo desses erros faz com que exista uma mudança nos sinais dos processos naturais do funcionamento do genoma, e isso acarreta em erros na maquinaria de genes de reparo.

Já a origem de CRC deve-se pela sucessão dos acúmulos de erros observados em adenomas, e a progressão desses erros nos genes responsáveis pelo reparo do genoma e sinalização de dano ao DNA fazem com as estruturas celulares saudáveis deem lugar a células mutadas em seus sistemas, e isso gera células cancerígenas, e lesões malignas em todo o tecido do colo e do reto.

Como esperado em IBD as respostas dos genes responsáveis pelas reações inflamatórias estão alteradas nesse tipo de fenótipo em relação a um tecido normal de colo. Como IBD é uma doença que causa lesões benignas, as respostas dos genes de reparo ao dano do DNA se mostram ativas, aparentemente normais e não alteradas significativamente nessa amostra, isto é indicio de que a via de reparo ao DNA ainda mantém a maquinaria da célula funcionando corretamente mesmo com lesões causadas por processos inflamatórios.

Em MSI não nota-se um padrão de comportamento das vias, mesmo havendo mudanças significativas de atividade em todas elas, exceto em apoptose. Isso

provavelmente se deve pelo fato da disfunção de todo o sistema de reparo, onde mudanças nos sinais e mutações pontuais dos genes, acumulam uma grande quantidade de erros, comprometendo assim todas as demais vias. A via do ciclo celular está muito alterada como mostra o Viacomplex®, indicando uma total disfunção estrutural nos processos realizados pela célula em MSI.

Uma forte convicção para esses indícios observados é o fato de apoptose não estar com atividade significativa, porém mostrar uma atividade diferenciada dos outros fenótipos. Observando a análise feita pelo Viacomplex®, a parte dos genes da rede que está alterada em apoptose é diferente dos genes que estão alterados em adenoma e CRC.

É importante ressaltar que todos os tipos de lesões, exibem algum grau de alteração na resposta inflamatória, alteração nos processos do ciclo celular, e todas respondem em algum grau de atividade alterada nos genes da via de apoptose. Também é notável uma disfunção nos sistemas de reparo do DNA no nível de dano, seja ele como origem, como causa, ou consequência das mutações observadas nos genes dessa via.

Esse panorama observado, nos mostra que mesmo havendo uma relação entre doenças, lesões e mutações que estão diretamente ligadas ao desenvolvimento do câncer de colorretal, cada um dos fenótipos possui um comportamento peculiar. No caso da progressão do adenoma para um fenótipo de câncer esporádico por sucessão de lesões acumulativas, o padrão das alterações se mostrou condizente com o modelo Fearon & Vogelstein (Fearon & Vogelstein, 1990) e os demais trabalhos que utilizaram esse modelo como base e ponto de partida, onde as mutações descritas pelo modelo causam alteração em toda a sinalização e nos processos de reparo, assim comprometendo o ciclo celular e apoptose, que acabam por degradar a estrutura normal das células nesse tecido. Porém é notável uma diferença de atividade dos genes entre dois fenótipos de câncer maligno, como observado contrastando os perfis de CRC e MSI. Isso sugere uma divergência entre duas causas em uma neoplasia (colorretal) que pode ser erroneamente estudado por possuir como resultado final um câncer aparentemente igual em ambas, ou seja, mesmo esses dois tipos de fenótipos ocasionando um cancro de colorretal, é errado enquadrá-los igualmente, pois como observado no presente estudo, possuem



causas e processo de desenvolvimento distinto entre si, e isso revela a variabilidade existente em tipos de câncer de colorretal.

## 6.2 CONSIDERAÇÕES ESPECÍFICAS

Como citado anteriormente, IBD é uma doença que causa reações inflamatórias da mucosa do intestino, atingindo também, colo e reto. Está diretamente envolvida com mutações em genes responsáveis pela resposta imunológica de controle da microbiota, ou com o erro na resposta em altas taxas de proliferação. Isso sugere que o padrão funcional de inflamação nessa rede esteja envolvida, nos erros de sinalização e de produção de células do sistema imunológico que atuam no reconhecimento e na proliferação de micróbios luminais. Esse quadro também sugere que lesões benignas consecutivas, estão envolvidas com o desenvolvimento do câncer de colorretal em pessoas acometidas com essa doença. Segundo os resultados encontrados, a expressão da via dos genes responsáveis pelo reparo do DNA se mostrou com funcionamento não diferencial e não significativo em relação a uma amostra de tecido normal, apenas com uma pequena variação notada na análise topográfica feita pelo software ©Viacomplex, mas descartada como significativa na análise estatística e GSEA, evidenciando que os sistemas de reparo e consequentemente do ciclo celular ainda mantém seu padrão de funcionamento nesse tipo de lesão. A provável causa para que isso ocorra, é o fato de se tratar de uma lesão benigna, onde neste caso, somente os processos inflamatórios estão com uma expressão diferencial enriquecida em relação ao controle, excluindo assim em primeira instância, a hipótese de que exista alterações significativas nos sinais e mecanismos de reparo nesse tipo de lesão. A profilaxia da doença muito provavelmente está envolvida na constante lesão do tecido por processos de inflamação, o que posteriormente pode causar o desenvolvimento de uma adenoma, e assim a doença seguir o padrão apresentado por Fearon & Vogelstein (Fearon & Vogelstein, 1990).

A evidência que a amostra de tecido com o padrão de lesão de IBD sugere, é um grau de diferença na via de inflamação da rede, se comparada à adenoma, CRC e MSI, assim como esperado para esse tipo de doença.

A progressão adenoma-carcinoma é caracterizada pela diferenciação de pólipos benignos em pólipos displásicos, já considerados lesões pré-malignas, classificados como adenomas, que podem se diferenciar em neoplasias malignas caso não haja a identificação e tratamento (Cotran et al., 2000). Sabe-se também que o acúmulo de erros nos genes responsáveis pelo reparo do DNA estão como um dos processos de evolução da lesão pré-maligna para o estado de carcinoma. No ensaio de adenoma deste estudo, a via dos genes de reparo se mostrou significativamente alterada, comprovando a hipótese citada anteriormente, onde os acúmulos de erros nestes genes, seriam responsáveis por modificar toda estrutura de funcionamento molecular da célula, gerando assim pólipos pré-malignos. Em CRC a atividade da via dos genes de reparo ao dano do DNA, já não se mostra significativa. Este resultado observado em nosso estudo, nos dá indícios de uma possível hipótese, em que os genes que desempenham o papel de reparo de adenoma não possuem o mesmo padrão de comportamento em CRC. É importante ressaltar, que os genes que possuem dupla atividade, nas vias de reparo e do ciclo celular estão significativos, evidenciando que as mudanças nesse tipo de fenótipo localizam-se nas partes da rede envolvidas nos sinais de reparo, mais especificamente nos genes que também desempenham atividade no processo do ciclo celular, reforçando que essas mudanças, modificam as conformações estruturais da célula neoplásica, o que gera o aspecto de massa tumoral no câncer. A via de apoptose está diferencialmente expressa em uma área específica da rede, como mostra a figura topográfica gerada pelo ©Viacomplex, o que não é um padrão normalmente observado em estudos anteriores em outros tecidos afetados pelo padrão de progressão neoplásica, e essa evidência é confirmada pela significância mostrada na análise de GSEA. Uma hipótese plausível para esse comportamento é o fato de haver uma grande desordem nos processos celulares, onde a célula teria sinais de morte celular programada, que não necessariamente estariam sendo executadas em um processo físico real na célula afetada, ou de que mesmo sofrendo apoptose, o crescimento do bolo tumoral pelas mutações observadas nas demais vias seria superior a esses sinais apoptóticos. Isso é reforçado pela significância observada nos processos inflamatórios em CRC, onde há diferença de atividade na expressão de um determinado número de genes mas não nos módulos com um todo.



A amostra de adenoma e de CRC foram próximas entre si, possuindo um grau de semelhança maior que os outros tipos de variações no tecido de colorretal desse estudo.

As regiões microssatélites são constituídos de pequenas repetições nucleotídicas e mutações, como por exemplo a duplicação dessas regiões. Causam erros estruturais, principalmente em regiões que possuem a codificação para a produção de proteínas funcionais no papel de reparo do DNA. A amostra de MSI utilizada nessa análise possui perfis de lesão maligna contrastadas com amostras de células estáveis em outros tecidos. Por possuir esse tipo de perfil que ocasiona mudanças estruturais, todo funcionamento das vias celulares nesse tipo de fenótipo deveria estar alterado significativamente.

Como citado, as evidências apresentadas pelos resultados, confirmam este indício, onde existe um aumento difuso em todas as vias da rede, exceto na via de apoptose, o que é esperado em tumores com características malignas.

O que chama atenção em MSI, é a diferença de expressão diferencial com relação às demais amostras de lesões nesse tipo de tecido, onde nota-se um aumento difuso e sem padrão definido na atividade em um ou mais módulos, uma inativação de apoptose, ressaltando que o perfil maligno de CRC da GSE 4183, possui atividade observada em apoptose, tanto na análise de atividade capturada pelo ©Viacomplex quanto pelo teste estatístico de GSEA. Mesmo com um padrão de ativação não concentrado nas vias, a significância é alta, e este aumento mostra que as causas geradas pelas características peculiares desse tipo de epidemiologia está diretamente ligada ao resultado final da doença.

De todas as amostras de tecido testadas nesse estudo o perfil observado em MSI é o que mais se difere em relação as demais, não somente em significância das vias, mas principalmente na quantidade de locais ativos na rede, mostrando um perfil de expressão totalmente diferente de um tecido estável de colorretal.

## 7. CONCLUSÃO

O câncer de colorretal é uma neoplasia que possui diferentes facetas, e apresenta uma forma de progressão e estabelecimento diferente dependendo das origens das lesões que acometem esse tecido. As vias presentes na reconstrução da rede de estabilidade do genoma, nos mostram diferença de expressão em todos os tipos de lesão que nos permite pontuar as seguintes conclusões:

- A reconstrução da rede de estabilidade do genoma e as etapas do *pipeline* apresentado mostraram-se um método conceitual eficaz para a avaliação de padrões de expressão em tecidos que estão alterados com lesões benignas, pré-malignas e malignas.
- A análise funcional de atividade de expressão gerada pelo ©Viacomplex foi confirmada pelos métodos estatísticos, onde a atividade funcional de expressão está demonstrada da mesma forma pela significância observada nas vias.
- Como esperado, o estudo evidencia que o câncer de colorretal possui um perfil de mutação distinto dependendo da lesão que acomete o tecido, onde caminhos diferentes levam aparentemente ao desenvolvimento de um cancro.
- A estabilidade dos mecanismos de reparo do DNA é imprescindível para o não desenvolvimento de neoplasias, e a mudança em sua funcionalidade e estrutura de processos e sinalização para as demais vias está diretamente ligado a profilaxia da doença.
- IBD, demonstrou um perfil de alteração nas vias de inflamação e uma mudança sutil nos genes de reparo ao dano do DNA, o que sugere a progressão para adenoma.

- Os genes de reparado do DNA estão alterados em adenoma, sugerindo que a mutação dos genes dessa via está diretamente ligado com o desenvolvimento da doença.
- CRC é uma provável consequência de acúmulos de mutações geradas na progressão de adenoma para uma lesão maligna.
- MSI possui um perfil displásico de atividade funcional na expressão das vias, exceto em apoptose, é o perfil mais alterado com relação a um tecido estável em todas as vias da rede de estabilidade do genoma.

## 8. PERSPECTIVAS

- Explorar e encontrar evidências mais sólidas para atividade de apoptose em CRC, e se atividade de expressão é relativa a uma alta taxa de transcritos que por alguma razão não desempenha um papel funcional.
- Aprofundar a análise funcional de MSI e encontrar um padrão funcional para o aumento difuso de comportamento observado na rede de estabilidade do genoma.
- Elucidar as causas de todas as vias da rede, estarem alteradas em MSI com exceção de apoptose.
- Colaboração na implementação do software ©Viacomplex como pacote para a linguagem R.

## 9. REFERÊNCIAS

- ALLTONEN, L. A.; PELTOMAKI, P.; LEACH, F. S.; SISTONEN, P.; PYLKKAENE, L.; MECKLIN, J. P.; JAERVINEN, H.; POWELL, S. M.; JEN, J.; HAMILTON, S. R.; PETERSON, G. M.; KINZLER, K. W.; VOLGELTEIN, B.; De la CHAPELLE, A. Clues to the pathogenesis of familial colorectal cancer. **Science**, v. 260, p. 812-816, 1993.
- ALANO, C. C.; YING, W.; SWANSON, R. A.; Poly(ADP-ribose) Polymerase-1-mediated Cell Death in Astrocytes Requires NAD<sup>+</sup> Depletion and Mitochondrial Permeability Transition. **Journal Of Biological Chemistry**, v. 279, n. 18, p.18895-18902, 2004.
- ALBERTSON, D. G. et al.; Chromosome aberrations in solid tumors. **Nature Genetics**, v. 34, n. 4, p.369-376, 2003.
- ANDERSON, Alexander R. A.; QUARANTA, Vito. Integrative mathematical oncology. **Nature Reviews Cancer**, v. 8, n. 3, p.227-234, 2008.
- BÄCKVALL, H. et al.; Genetic tumor archeology: microdissection and genetic heterogeneity in squamous and basal cell carcinoma. **Mutation Research/fundamental And Molecular Mechanisms Of Mutagenesis**, v. 571, n. 1-2, p.65-79, 2005.
- BARABÁSI, A. L.; OLTVAI, Z. N.; Network biology: understanding the cell's functional organization. **Nature Reviews Genetics**, v. 5, n. 2, p.101-113, 2004.
- BENNETT, J. M. et al.; Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Co-operative Group. **British Journal Of Haematology**, v. 33, n. 4, p.451-458, 1976.
- BIRD, A. et al. DNA methylation patterns and epigenetic memory. **Genes & Development**, v. 16, n. 1, p.6-21, 2002.
- BIRNEY, Ewan et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. **Nature**, v. 447, n. 7146, p.799-816, 2007.
- CASTRO, M. A. A. et al. Evolutionary origins of human apoptosis and genome-stability gene networks. **Nucleic Acids Research**, v. 36, n. 19, p.6269-6283, 2008.
- CASTRO, M. A. A. et al. Impaired expression of NER gene network in sporadic solid tumors. **Nucleic Acids Research**, v. 35, n. 6, p.1859-1867, 2007.
- CASTRO, M. A. A. et al. ViaComplex: software for landscape analysis of gene expression networks in genomic context. **Bioinformatics**, v. 25, n. 11, p.1468-1469, 2009.

CASTRO, M. A. A.; *Estudo da diversidade tumoral: E desenvolvimento de ferramentas de bioinformática para análise citogenética e molecular de neoplasias sólidas*. 2009. 163 f. Tese (Doutorado em Bioquímica) – Instituto de Ciências Básicas da Saúde, Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre. 2009.

CEOL, Craig J et al. APC and colon cancer: two hits for one. **Nature Medicine**, v. 13, n. 11, p.1286-1287, 2007.

CHEN, Guoli et al. Molecular Diagnostics for Precision Medicine in Colorectal Cancer: Current Status and Future Perspective. **Biomed Research International**, v. 2016, p.1-12, 2016.

COTRAN, R.; COLLINS, K. V.; ROBBINS, T. **Robbins: Patologia Estrutural e Funcional**. Rio de Janeiro: Guanabara Koogan, 2000.

CRESPI, B.; SUMMERS, K.; Evolutionary biology of cancer. **Trends In Ecology & Evolution**, v. 20, n. 10, p.545-552, 2005.

DANIAL, N.N.; KORSMEYER, S.J.; Cell death: Critical control points. **Cell**, v. 116, p.205–219, 2004.

DEBRAWERE, H.; GENDREL, C. G.; LECHAT, S.; DUTREIX, M. Differences and similarities between several tandem repeat sequences: minisatellites and microsatellites. **Biochimie**, v.79, p.577-86, 1997.

DI MEGLIO P., et al. The IL23R R381Q gene variant protects against immune-mediated diseases by impairing IL-23-induced Th17 effector response in humans. **PLoS ONE**, 6:e17160, 2011.

DUCKETT, D. R. et al. HMutSalpha - and hMutLalpha -dependent phosphorylation of p53 in response to DNA methylator damage. **Proceedings Of The National Academy Of Sciences**, v. 96, n. 22, p.12384-12388, 26 out. 1999.

EISENBERG, D.; et al. Protein function in the post-genomic era. **Nature**, v. 405, n. 6788, p.823-826, 2000.

ENCODE Project Consortium. Birney E, Stamatoyannopoulos E, Dutta A. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. **Nature**, v.447, n.7146, p. 799–816, 2007.

FEARON, Eric R.; VOGELSTEIN, Bert. A genetic model for colorectal tumorigenesis. **Cell**, v. 61, n. 5, p.759-767, 1990.

FODDE, Riccardo; SMITS, Ron; CLEVERS, Hans. APC, Signal transduction and genetic instability in colorectal cancer. **Nature Reviews Cancer**, v. 1, n. 1, p.55-67, 2001

FORBES, S. A. et al. The catalogue of somatic mutations in cancer (COSMIC). **Current Protocols in Human Genetics**, v. Chapter 10, p. Unit2008.

FRANCESCHINI, A. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. **Nucleic Acids Research**, D808-15, 2013.

GALAMB, O. et al.; Inflammation, adenoma and cancer: objective classification of colon biopsy specimens with gene expression signature. **Disease Markers**, v.25, n.1, p.1-16, 2008.

GAFA, Roberta et al. Sporadic colorectal adenocarcinomas with high-frequency microsatellite instability. **Cancer**, v. 89, n. 10, p.2025-2037, 2000.

GLOCKNER, S. et al.; Marked intratumoral heterogeneity of c-myc and cyclinD1 but not of c-erbB2 amplification in breast cancer. **Laboratory Investigation**. v.82, n. 10, p.1419-1426, 2002.

Guerra M.R.; Gallo C.V.M.; Mendonça G.A.S. Risco de câncer no Brasil: tendências e estudos epidemiológicos mais recentes. **Revista Brasileira de Cancerologia**; v.51(3), p. 227-234, 2005.

HAHN, W. C.; WEINBERG, R. A.; Rules for making human tumor cells. **New England Journal of Medicine**, v.347, n.20, p.1593-1603, 2002.

HARADA K. et al.; Intratumoral Cytogenetic Heterogeneity Detected by Comparative Genomic Hybridization and Laser Scanning Cytometry in Human Gliomas. **Cancer Research**, v. 58, n. 20, p.4694-4700, 1998.

HARRIS, N. L. et al.; A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma study group. **Current Diagnostic Pathology**, v. 2, n. 1, p.58-59, 1995.

HARRIS, N. L. et al.; World Health Organization Classification of Neoplastic Diseases of the Hematopoietic and Lymphoid Tissues: Report of the Clinical Advisory Committee Meeting Airlie House, Virginia, November 1997. **Journal Clinical Oncology**, v. 17, n. 12, p.3835-3849, 1999.

HENG, H. H. et al.; Clonal and non-clonal chromosome aberrations and genome variation and aberration. **Génome**, v. 49, n. 3, p.195-204, 2006.

HIPFNER, D. R.; COHEN, S. M.; Connecting proliferation and apoptosis in development and disease. **Nature Reviews Molecular Cell Biology**, v. 5, n. 10 p.805-815, 2004.

HUMPHRIES, Adam; WRIGHT, Nicholas A.. Colonic crypt organization and tumorigenesis. **Nature Reviews Cancer**, v. 8, n. 6, p.415-424, 2008.

ICD-O-1. **International Classification of Disease for Oncology**. 1.ed. Geneva: World Health Organization, 1976.

IKENAGA, Masakazu et al. Use of microsatellite analysis in young patients with colorectal cancer to identify those with hereditary nonpolyposis colorectal cancer. **Journal Of Surgical Oncology**, v. 79, n. 3, p.157-165, 2002.

INCA Instituto Nacional do Câncer. O que é câncer. Disponível em <[http://www1.inca.gov.br/conteudo\\_view.asp?id=322](http://www1.inca.gov.br/conteudo_view.asp?id=322)> Acesso 09 de julho de 2016.

IONOV, Y.; PEINADO, M. A.; MALKHOSYAN, S.; SHIBATA, D.; PERUCHO, M. Ubiquitous somatic mutations in simple repetitive sequences reveal a new mechanism for colonic carcinogenesis. **Nature**, v. 363, p. 558-576, 1993.

JEFFORD, Charles Edward et al. Mechanisms of chromosome instability in cancers. **Critical Reviews In Oncology/hematology**, v. 59, n. 1, p.1-14, 2006.

JONES, S. et al. Comparative lesion sequencing provides insights into tumor evolution. **Proceedings Of The National Academy Of Sciences**, v. 105, n. 11, p.4283-4288, 2008.

KANEHISA, M.; GOTO, S.; HATTORI, M.; AOKI-KINOSHITA, K. F.; ITOH, M.; KAWASHIMA, S.; KATAYAMA, T.; ARAKI, M.; HIRAKAWA, M.; From genomics to chemical genomics: new developments in KEGG. **Nucleic Acids Research**, v. 34, D354–D357, 2006.

KEEDWELL, E.; NARAYANAN, A. **Intelligent bioinformatics - The Application of Artificial Intelligence Techniques to Bioinformatics Problems**, England, Wiley, 2005.

KHOR, Bernard; GARDET, Agnès; XAVIER, Ramnik J.. Genetics and pathogenesis of inflammatory bowel disease. **Nature**, v. 474, n. 7351, p.307-317, 2011.

KINZLER, Kenneth W et al. Lessons from Hereditary Colorectal Cancer. **Cell**, v. 87, n. 2, p.159-170, 1996.

KUFE, D. et al. Cardinal manifestations of cancer. In: D. KUFE; R. BAST; W. HAIT; W. HONG; R. POLLACK; R. WEICHSELBAUM; J. HOLLAND; E. FREI (Eds). **Cancer Medicine**. 2005. p. ISBN 1-55009-307-X.

KUMAR, Vinay; ABBAS, Abul K.; ASTER, Jon C.: Robbins Patologia Básica. 9. ed. Rio de Janeiro: Elsevier, 2013. 928 p.

LAM, Alfred King-yin et al. Synchronous colorectal cancer: Clinical, pathological and molecular implications. **World Journal Of Gastroenterology**, v. 20, n. 22, p.6815-6820, 2014.

LETTRE, G.; HENGARTNER, M.O.; Developmental apoptosis in *C. elegans*: a complex CEDnario. **Nature Reviews Molecular Cell Biology**, v. 7, n. 2, p.97–108, 2006.



LEVINSON, G.; GUTMAN, G. A. High frequencies of short frameshifts in poly CA/TG tandem repeats done by bacteriophage M13 escherichia coli K12. **Nucleic Acids Research**, v.15, p.5323-5338, 1987.

LODISH, Steinberg Biosciences, G., Kim, P., & Lippincott-Schwartz, J. **Molecular Cell Biology 5<sup>th</sup> ed**, New York: W.H. Freeman and Co. 2003.

LOSI, L. et al.; Evolution of intratumoral genetic heterogeneity during colorectal cancer progression. **Carcinogenesis**, v. 26, n. 5, p.916-922, 2004.

MEMORIAL UNIVERSITY. MUN, Canadá, 2011. Disponível em: <<http://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-21/2110.jpg>>. Acesso em: 25 de julho de 2016.

MARRA, G., BOLAND, R. DNA repair and colorectal cancer. **Gastroenterol. Clin. North Am.**, v.25, p.755-772, 1996.

MITELMAN F.; Recurrent chromosomal aberrations in cancer. **Mutation Research**, v.462, n. 2-3, p.247-253, 2000.

MITELMAN, F. et al.; Clinical significance of cytogenetic findings in solid tumors. **Cancer Genetics And Cytogenetics**, v. 95, n. 1, p.1-8, 1997.

MOMOZAWA, Y., et al. Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. **Nature Genetics**, v. 43, n. 1, p.43-47, 2010.

NILBERT, M. et al. Microsatellite instability is rare in rectal carcinomas and signifies hereditary cancer. **European Journal Of Cancer**, v. 35, n. 6, p.942-945, 1999.

ORNDAL, C. et al. Cytogenetic evolution in primary tumors, local recurrences, and pulmonary metastases of two soft tissue sarcomas. **Clinical & Experimental Metastasis**, v. 11, n. 5, p.401-408, 1993.

PAPIN, J. A.; et al. Reconstruction of cellular signalling networks and analysis of their properties. **Nature Reviews Molecular Cell Biology**, v. 6, n. 2, p.99-111, 2005.

PSIQWEB. Locais de manifestação das doenças inflamatórias intestinais, 2005. Disponível em: <[http://www.psiqweb.med.br/site/FBArquivos%5CNO%5C00000047%5C00000047\\_84.jpg](http://www.psiqweb.med.br/site/FBArquivos%5CNO%5C00000047%5C00000047_84.jpg)>. Acesso em: 27 de julho de 2016.

RAPPAPORT, H. et al.; Report of the Committee on Histopathological Criteria Contributing to Staging of Hodgkin's Disease. **Cancer Research**, v. 31, n. 11, p. 1864-1865, 1971. Disponível em: PMID:5121694.

RYBARCZYK-FILHO, J. L. et al. Towards a genome-wide transcriptogram: the *Saccharomyces cerevisiae* case. **Nucleic Acids Research**, v. 39, n. 8, p.3005-3016, 2011.

SALETTI, P.; EDWIN, I. D.; PACK, K.; CAVALLI, F.; ATKIN, W. S. Microsatellite instability: Application in hereditary non-polyposis colorectal cancer. **Annals of Oncology**, v.12, p.151-160, 2001.

SENGUPTA, S.; HARRIS, C. C.; P53: traffic cop at the crossroads of DNA repair and recombination. **Nature Reviews Molecular Cell Biology**, v. 6, n. 1, p.44-55, 2005.

SIEGEL, Rebecca et al. Cancer statistics, 2011. **Ca: A Cancer Journal for Clinicians**, v. 61, n. 4, p.212-236, 2011.

SMYTH G. K., Limma: linear models for microarray data. In: 'Bioinformatics and Computational Biology Solutions using R and Bioconductor', **Springer**, New York, p. 397-420, 2005.

STAROSTIK, Petr et al. Diagnosis of microsatellite instability-positive colorectal cancer. **Expert Review Of Molecular Diagnostics**, v. 1, n. 1, p.71-80, 2001.

STRAND, M.; PROLLA, T. A.; LISKAY, R. M.; PETES, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. **Nature**, v. 365, p. 274-276, 1993.

THIBODEAU, S. N.; BREN, G.; SHAID, D. Microsatellite instability in cancer of the proximal colon. **Science**, v. 260, p. 816-819, 1993.

UNM Biology Undergraduate, University of New Mexico, 2002. Disponível em: <[http://biology.unm.edu/ccouncil/Biology\\_124/Images/RNAtranslation.jpeg](http://biology.unm.edu/ccouncil/Biology_124/Images/RNAtranslation.jpeg)>. Acesso em: 26 de julho de 2016.

Waddington C.H. The epigenotype. **Endeavour**. 1942;1:18–20

Wang X, Terfve C, Rose JC and Markowitz F. "HTSanalyzeR: an R/Bioconductor package for integrated network analysis of high-throughput screens." **Bioinformatics**, v.27, n.6, p. 879, 2011.

WATANABE T.; et al.; Distal colorectal cancers with microsatellite instability (MSI) display distinct gene expression profiles that are different from proximal MSI cancers. **Cancer Research**, v.66, n.20, p.9804-9808, 2006.

WATSON, J. D.; BAKER, T. A.; BELL, S. P.; GANN, A.; LEVINE, M.; LOSICK, R.: **Molecular Biology of the Gene**, 7 ed, 2014.

WEBER, J. L. & MAY, P. E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. **Am. J. Hum. Genet.**, v.44, p.388-96, 1989.

WIKICIENCIAS. Ciclo Celular, 2012. Disponível em: <<http://wikiciencias.casadasciencias.org/wiki/images/4/4f/Ciclocelular.png>> Acesso em: 15 de junho 2016

YAN, B. et al. A unique role of the DNA fragmentation factor in maintaining genomic stability. **Proceedings Of The National Academy Of Sciences**, v. 103, n. 5, p.1504-1509, 2006.

ZHIVOTOVSKY, Boris; KROEMER, Guido. Apoptosis and genomic instability. **Nature Reviews Molecular Cell Biology**, v. 5, n. 9, p.752-762, 2004.

**ANEXO - LISTA DE GENES RETIRADOS DO ONTOCANCRO 2.0 PARA CONFEÇÃO DA REDE DOS PROCESSOS DE ESTABILIDADE DO GENOMA.**

Approved_Symbol	Apoptosis	Inflammation	DNA_Damage_Response	Cell_Cycle	Apoptosis_Inflammation	DNA_Damage_Cell_Cycle	Intersection
ABL1	0	0	1	1	0	1	0
ACIN1	1	0	0	0	0	0	0
ACOT13	0	0	0	0	0	0	0
ACR	0	0	0	0	0	0	0
ACTA1	1	0	0	0	0	0	0
ACTG1	0	0	0	0	0	0	0
ACTN4	0	0	0	0	0	0	0
ACTR1A	0	0	0	1	0	0	0
ADAM17	1	0	0	0	0	0	0
ADCYAP1R1	0	0	0	0	0	0	0
ADD1	1	0	0	0	0	0	0
ADRBK1	0	1	0	0	0	0	0
AHCTF1	0	0	0	1	0	0	0
AIFM1	1	0	0	0	0	0	0
AKAP9	0	0	0	1	0	0	0
AKT1	1	1	0	0	1	0	0
AKT1S1	0	1	0	0	0	0	0
AKT2	1	0	0	0	0	0	0
AKT3	1	0	0	0	0	0	0
ALMS1	0	0	0	1	0	0	0
ALOX5	0	1	0	0	0	0	0
ANAPC1	0	0	0	1	0	0	0
ANAPC10	0	0	0	1	0	0	0
ANAPC11	0	0	0	1	0	0	0

ANAPC2	0	0	0	1	0	0	0
ANAPC4	0	0	0	1	0	0	0
ANAPC5	0	0	0	1	0	0	0
ANAPC7	0	0	0	1	0	0	0
ANP32A	1	0	0	0	0	0	0
APAF1	1	0	0	0	0	0	0
APC	1	1	0	0	1	0	0
APEX1	1	0	1	0	0	0	0
APEX2	0	0	1	0	0	0	0
APITD1	0	0	0	1	0	0	0
APP	1	0	0	0	0	0	0
APPL1	1	0	0	0	0	0	0
ARHGAP10	1	0	0	0	0	0	0
ARHGDIB	1	0	0	0	0	0	0
ARRB1	0	1	0	0	0	0	0
ARRB2	0	1	0	0	0	0	0
ATF1	0	1	0	0	0	0	0
ATF2	0	1	0	0	0	0	0
ATF6	0	1	0	0	0	0	0
ATG13	0	1	0	0	0	0	0
ATM	1	1	1	1	1	1	1
ATP5J	0	0	0	0	0	0	0
ATR	0	1	1	1	0	1	1
ATRIP	0	1	1	1	0	1	1
AURKA	0	0	0	1	0	0	0
AURKB	0	0	0	1	0	0	0
AXIN1	0	1	0	0	0	0	0
AXIN2	0	0	0	0	0	0	0
AZI1	0	0	0	1	0	0	0
B9D2	0	0	0	1	0	0	0
BAD	1	0	0	0	0	0	0
BAG4	1	0	0	0	0	0	0
BAK1	1	0	0	0	0	0	0
BARD1	0	0	0	0	0	0	0
BAX	1	1	0	0	1	0	0
BBC3	1	0	0	0	0	0	0
BCAP31	1	0	0	0	0	0	0
BCL10	0	1	0	0	0	0	0
BCL2	1	1	0	0	1	0	0
BCL2L1	1	1	0	0	1	0	0
BCL2L11	1	0	0	0	0	0	0
BCL3	0	1	0	0	0	0	0
BID	1	0	0	0	0	0	0
BIRC2	1	1	0	0	1	0	0

BIRC3	1	1	0	0	1	0	0
BIRC5	0	0	0	1	0	0	0
BLK	0	1	0	0	0	0	0
BLM	0	0	1	0	0	0	0
BLVRA	0	1	0	0	0	0	0
BLVRB	0	1	0	0	0	0	0
BMF	1	0	0	0	0	0	0
BMI1	0	0	0	0	0	0	0
BMX	1	0	0	0	0	0	0
BNIP3	0	1	0	0	0	0	0
BRAF	0	1	0	0	0	0	0
BRCA1	0	1	1	1	0	1	1
BRCA2	0	0	1	0	0	0	0
BRIP1	0	0	1	0	0	0	0
BTRC	0	0	0	1	0	0	0
BUB1	0	0	0	1	0	0	0
BUB1B	0	0	0	1	0	0	0
BUB3	0	0	0	1	0	0	0
C17orf70	0	0	1	0	0	0	0
C19orf40	0	0	1	0	0	0	0
CACYBP	0	0	0	0	0	0	0
CALR	0	0	0	0	0	0	0
CAPN1	1	0	0	0	0	0	0
CAPN2	1	0	0	0	0	0	0
CASC5	0	0	0	1	0	0	0
CASP1	1	0	0	0	0	0	0
CASP10	1	0	0	0	0	0	0
CASP2	1	0	0	0	0	0	0
CASP3	1	0	0	0	0	0	0
CASP4	1	0	0	0	0	0	0
CASP6	1	0	0	0	0	0	0
CASP7	1	0	0	0	0	0	0
CASP8	1	1	0	0	1	0	0
CASP9	1	0	0	0	0	0	0
CAV1	1	0	0	0	0	0	0
CBL	0	1	0	0	0	0	0
CBX5	0	0	0	0	0	0	0
CCDC99	0	0	0	1	0	0	0
CCNA1	0	0	0	1	0	0	0
CCNA2	0	0	1	1	0	1	0
CCNB1	0	0	0	1	0	0	0
CCNB2	0	0	0	1	0	0	0
CCNC	0	0	0	0	0	0	0
CCND1	0	1	0	1	0	0	1

CCND2	0	1	0	1	0	0	1
CCND3	0	0	0	1	0	0	0
CCNE1	0	1	0	1	0	0	1
CCNE2	0	1	0	1	0	0	1
CCNH	0	0	1	1	0	1	0
CCNO	0	0	1	0	0	0	0
CDC14A	0	0	0	1	0	0	0
CDC16	0	0	0	1	0	0	0
CDC2	0	0	0	1	0	0	0
CDC20	0	0	0	1	0	0	0
CDC23	0	0	0	1	0	0	0
CDC25A	0	1	0	1	0	0	1
CDC25B	0	1	0	1	0	0	1
CDC25C	0	0	1	1	0	1	0
CDC26	0	0	0	1	0	0	0
CDC27	0	0	0	1	0	0	0
CDC34	0	0	0	1	0	0	0
CDC45L	0	1	0	1	0	0	1
CDC6	0	0	1	1	0	1	0
CDC7	0	0	0	1	0	0	0
CDCA4	0	0	0	0	0	0	0
CDCA7	0	0	0	0	0	0	0
CDCA8	0	0	0	1	0	0	0
CDH1	1	0	0	0	0	0	0
CDK11A	0	0	0	1	0	0	0
CDK11B	0	0	0	1	0	0	0
CDK2	0	1	1	1	0	1	1
CDK4	0	1	0	1	0	0	1
CDK5RAP2	0	0	0	1	0	0	0
CDK6	0	1	0	1	0	0	1
CDK7	0	0	1	1	0	1	0
CDKN1A	0	1	0	1	0	0	1
CDKN1B	0	0	0	1	0	0	0
CDKN2A	0	1	0	1	0	0	1
CDKN2B	0	0	0	1	0	0	0
CDKN2C	0	0	0	1	0	0	0
CDKN2D	0	0	0	1	0	0	0
CDT1	0	0	0	1	0	0	0
CEBPA	0	1	0	0	0	0	0
CEBPB	0	1	0	0	0	0	0
CEBPD	0	1	0	0	0	0	0
CENPA	0	0	0	1	0	0	0
CENPC1	0	0	0	1	0	0	0
CENPE	0	0	0	1	0	0	0

CENPF	0	0	0	1	0	0	0
CENPH	0	0	0	1	0	0	0
CENPI	0	0	0	1	0	0	0
CENPJ	0	0	0	1	0	0	0
CENPK	0	0	0	1	0	0	0
CENPL	0	0	0	1	0	0	0
CENPM	0	0	0	1	0	0	0
CENPN	0	0	0	1	0	0	0
CENPO	0	0	0	1	0	0	0
CENPP	0	0	0	1	0	0	0
CENPQ	0	0	0	1	0	0	0
CENPT	0	0	0	1	0	0	0
CEP110	0	0	0	1	0	0	0
CEP135	0	0	0	1	0	0	0
CEP152	0	0	0	1	0	0	0
CEP164	0	0	1	1	0	1	0
CEP192	0	0	0	1	0	0	0
CEP250	0	0	0	1	0	0	0
CEP290	0	0	0	1	0	0	0
CEP57	0	0	0	1	0	0	0
CEP63	0	0	0	1	0	0	0
CEP70	0	0	0	1	0	0	0
CEP72	0	0	0	1	0	0	0
CEP76	0	0	0	1	0	0	0
CEP78	0	0	0	1	0	0	0
CETN2	0	0	1	1	0	1	0
CFL2	1	0	0	0	0	0	0
CFLAR	1	0	0	0	0	0	0
CHEK1	0	1	1	1	0	0	1
CHEK2	0	1	1	1	0	0	1
CHUK	1	1	0	0	1	0	0
CKAP5	0	0	0	1	0	0	0
CKS1B	0	0	0	1	0	0	0
CLASP1	0	0	0	1	0	0	0
CLASP2	0	0	0	1	0	0	0
CLIP1	0	1	0	1	0	0	1
CLSPN	1	1	1	1	1	1	1
COX4I1	0	0	0	0	0	0	0
CRADD	1	0	0	0	0	0	0
CREB1	0	1	0	0	0	0	0
CREBBP	1	0	0	0	0	0	0
CRP	0	1	0	0	0	0	0
CRTC1	0	1	0	0	0	0	0
CRTC2	0	1	0	0	0	0	0



CSDA	0	0	0	0	0	0	0
CSF2RB	1	0	0	0	0	0	0
CSNK1D	0	0	0	1	0	0	0
CSNK1E	0	0	0	1	0	0	0
CSNK2A1	0	1	0	0	0	0	0
CSNK2A2	0	1	0	0	0	0	0
CSTF1	0	0	0	0	0	0	0
CSTF3	0	0	0	0	0	0	0
CTNNB1	1	1	0	0	1	0	0
CUL1	0	0	0	1	0	0	0
CUL4A	0	0	1	0	0	0	0
CUL4B	0	0	1	0	0	0	0
CXCL8	0	1	0	0	0	0	0
CXCR1	0	1	0	0	0	0	0
CXCR2	0	1	0	0	0	0	0
CYCS	1	0	0	0	0	0	0
CYLD	1	1	0	0	1	0	0
DAPK1	1	0	0	0	0	0	0
DAPK2	1	0	0	0	0	0	0
DAPK3	1	0	0	0	0	0	0
DBF4	0	0	0	1	0	0	0
DBF4B	0	0	0	0	0	0	0
DBNL	1	0	0	0	0	0	0
DCC	1	0	0	0	0	0	0
DCLRE1C	0	0	1	0	0	0	0
DCTN1	0	0	0	1	0	0	0
DCTN2	0	0	0	1	0	0	0
DCTN3	0	0	0	1	0	0	0
DDB1	0	0	1	0	0	0	0
DDB2	0	0	1	0	0	0	0
DDIT3	0	1	0	0	0	0	0
DDIT4	0	1	0	0	0	0	0
DDX11	0	0	0	0	0	0	0
DDX5	0	0	0	0	0	0	0
DDX58	0	1	0	0	0	0	0
DEFB4A	0	1	0	0	0	0	0
DEPTOR	0	1	0	0	0	0	0
DFFA	1	0	0	0	0	0	0
DFFB	1	0	0	0	0	0	0
DHFR	0	0	0	1	0	0	0
DIABLO	1	0	0	0	0	0	0
DIRAS3	0	0	0	0	0	0	0
DMC1	0	0	1	0	0	0	0
DNA2	0	0	0	1	0	0	0

DNM1	0	1	0	0	0	0	0
DNM1L	1	0	0	0	0	0	0
DNMT1	0	1	0	0	0	0	0
DNMT3B	0	1	0	0	0	0	0
DNTT	0	0	1	0	0	0	0
DOCK2	0	1	0	0	0	0	0
DSG1	1	0	0	0	0	0	0
DSG2	1	0	0	0	0	0	0
DSG3	1	0	0	0	0	0	0
DSN1	0	0	0	1	0	0	0
DSP	1	0	0	0	0	0	0
DUSP1	0	1	0	0	0	0	0
DUSP10	0	1	0	0	0	0	0
DUSP16	0	1	0	0	0	0	0
DUSP8	0	1	0	0	0	0	0
DVL1	0	1	0	0	0	0	0
DYNC1H1	0	0	0	1	0	0	0
DYNC1I2	0	0	0	1	0	0	0
DYNLL1	1	0	0	1	0	0	0
DYNLL2	1	0	0	0	0	0	0
E2F1	1	1	0	1	0	0	1
E2F2	0	1	0	1	0	0	0
E2F3	0	0	0	1	0	0	0
E2F4	0	0	0	1	0	0	0
E2F5	0	0	0	1	0	0	0
E2F6	0	0	0	0	0	0	0
E2F7	0	0	0	0	0	0	0
E2F8	0	0	0	0	0	0	0
ECE1	0	0	0	0	0	0	0
EED	0	0	0	0	0	0	0
EEF2	0	1	0	0	0	0	0
EEF2K	0	1	0	0	0	0	0
EIF1AX	0	0	0	0	0	0	0
EIF2AK2	0	0	0	0	0	0	0
EIF2S1	0	0	0	0	0	0	0
EIF2S2	0	0	0	0	0	0	0
EIF2S3	0	0	0	0	0	0	0
EIF3A	0	0	0	0	0	0	0
EIF3B	0	0	0	0	0	0	0
EIF3C	0	0	0	0	0	0	0
EIF3CL	0	0	0	0	0	0	0
EIF3D	0	0	0	0	0	0	0
EIF3E	0	0	0	0	0	0	0
EIF3F	0	0	0	0	0	0	0

EIF3G	0	0	0	0	0	0	0
EIF3H	0	0	0	0	0	0	0
EIF3I	0	0	0	0	0	0	0
EIF3J	0	0	0	0	0	0	0
EIF3K	0	0	0	0	0	0	0
EIF4A1	0	1	0	0	0	0	0
EIF4A2	0	0	0	0	0	0	0
EIF4A3	0	0	0	0	0	0	0
EIF4B	0	1	0	0	0	0	0
EIF4E	0	1	0	0	0	0	0
EIF4E2	0	0	0	0	0	0	0
EIF4EBP1	0	1	0	0	0	0	0
EIF4G1	0	0	0	0	0	0	0
EIF4G2	0	0	0	0	0	0	0
EIF4G3	0	0	0	0	0	0	0
EIF4H	0	0	0	0	0	0	0
EIF5	0	0	0	0	0	0	0
ELMO1	0	1	0	0	0	0	0
EME1	0	0	1	0	0	0	0
EME2	0	0	1	0	0	0	0
ENDOG	1	0	0	0	0	0	0
EP300	0	0	0	1	0	0	0
ERC1	0	1	0	0	0	0	0
ERCC1	0	0	1	0	0	0	0
ERCC2	0	0	1	0	0	0	0
ERCC3	0	0	1	0	0	0	0
ERCC4	0	0	1	0	0	0	0
ERCC5	0	0	1	0	0	0	0
ERCC6	0	0	1	0	0	0	0
ERCC6L	0	0	0	1	0	0	0
ERCC8	0	0	1	0	0	0	0
ESR1	0	1	0	0	0	0	0
ETV1	0	1	0	0	0	0	0
EXO1	0	0	1	0	0	0	0
EXOSC8	0	0	0	0	0	0	0
EZH2	0	0	0	0	0	0	0
FADD	1	1	0	0	1	0	0
FAM175A	0	1	0	0	0	0	0
FANCA	0	0	1	0	0	0	0
FANCB	0	0	1	0	0	0	0
FANCC	0	0	1	0	0	0	0
FANCD2	0	0	1	0	0	0	0
FANCE	0	0	1	0	0	0	0
FANCF	0	0	1	0	0	0	0

FANCG	0	0	1	0	0	0	0
FANCI	0	0	1	0	0	0	0
FANCL	0	0	1	0	0	0	0
FANCM	0	0	1	0	0	0	0
FAS	1	0	0	0	0	0	0
FASLG	1	0	0	0	0	0	0
FAU	0	0	0	0	0	0	0
FBXO5	0	0	0	1	0	0	0
FBXW11	0	1	1	0	0	0	1
FEN1	0	0	1	1	0	0	0
FGFR1	0	0	0	0	0	0	0
FGFR1OP	0	0	0	1	0	0	0
FGFR2	0	0	0	0	0	0	0
FGG	0	1	0	0	0	0	0
FGR	0	1	0	0	0	0	0
FHIT	0	0	0	0	0	0	0
FNTA	1	0	0	0	0	0	0
FOS	0	1	0	0	0	0	0
FOXO1	0	1	0	0	0	0	0
FOXO3	0	0	0	0	0	0	0
FSHR	0	0	0	0	0	0	0
FYN	0	1	0	0	0	0	0
FZR1	0	0	0	1	0	0	0
GAB1	0	1	0	0	0	0	0
GAB2	0	1	0	0	0	0	0
GADD45A	0	0	1	1	0	1	0
GAS2	1	0	0	0	0	0	0
GDI1	0	1	0	0	0	0	0
GIN51	0	0	0	1	0	0	0
GIN52	0	0	0	1	0	0	0
GIN54	0	0	0	1	0	0	0
GMNN	0	0	0	1	0	0	0
GNA14	0	1	0	0	0	0	0
GNA15	0	1	0	0	0	0	0
GNAI2	0	1	0	0	0	0	0
GNB1	0	1	0	0	0	0	0
GNB2L1	1	0	0	0	0	0	0
GNG2	0	1	0	0	0	0	0
GORASP1	0	0	0	1	0	0	0
GRB2	0	1	0	0	0	0	0
GSK3B	0	1	0	1	0	0	1
GSN	1	0	0	0	0	0	0
GTF2H1	0	0	1	0	0	0	0
GTF2H2	0	0	1	0	0	0	0

GTF2H3	0	0	1	0	0	0	0
GTF2H4	0	0	1	0	0	0	0
GTF2H5	0	0	1	0	0	0	0
GZMA	1	0	0	0	0	0	0
GZMB	1	0	0	0	0	0	0
H19	0	0	0	0	0	0	0
H1FO	1	0	0	0	0	0	0
H2AFX	0	1	1	0	0	0	1
HAUS2	0	0	0	1	0	0	0
HBP1	0	1	0	0	0	0	0
HCK	0	1	0	0	0	0	0
HDAC1	0	0	0	1	0	0	0
HIST1H1A	1	0	0	0	0	0	0
HIST1H1B	1	0	0	0	0	0	0
HIST1H1C	1	0	0	0	0	0	0
HIST1H1D	1	0	0	0	0	0	0
HIST1H1E	1	0	0	0	0	0	0
HIST1H2AC	0	1	0	0	0	0	0
HMGB1	1	0	0	0	0	0	0
HMGB1L1	0	0	1	0	0	0	0
HMGB1L10	0	0	1	0	0	0	0
HMGB2	1	0	1	0	0	0	1
HMOX1	0	1	0	0	0	0	0
HOXA3	0	0	0	0	0	0	0
HRAS	0	1	0	0	0	0	0
HRK	0	0	0	0	0	0	0
HSP90AA1	0	0	0	1	0	0	0
HSP90B1	0	1	0	0	0	0	0
HSPB1	0	1	0	0	0	0	0
HUS1	0	0	1	1	0	1	0
IKKB	1	1	0	0	1	0	0
IKBK	1	1	0	0	1	0	0
IL10	0	1	0	0	0	0	0
IL10RA	0	1	0	0	0	0	0
IL10RB	0	1	0	0	0	0	0
IL1A	1	1	0	0	1	0	0
IL1B	1	0	0	0	0	0	0
IL1R1	1	1	0	0	1	0	0
IL1RAP	1	0	0	0	0	0	0
IL3	1	0	0	0	0	0	0
IL3RA	1	0	0	0	0	0	0
IL6	0	1	0	0	0	0	0
IL6R	0	1	0	0	0	0	0
IL6ST	0	1	0	0	0	0	0

INCENP	0	0	0	1	0	0	0
IRAK1	1	1	0	0	1	0	0
IRAK2	1	0	0	0	0	0	0
IRAK3	1	0	0	0	0	0	0
IRAK4	1	0	0	0	0	0	0
IRF1	0	1	0	0	0	0	0
IRF6	0	1	0	0	0	0	0
IRS1	0	1	0	0	0	0	0
ISYNA1	0	0	0	0	0	0	0
ITGB3BP	0	0	0	1	0	0	0
JAK1	0	1	0	0	0	0	0
JAK2	0	1	0	0	0	0	0
JMY	0	0	0	0	0	0	0
JPH2	0	0	0	0	0	0	0
JUN	0	1	1	0	0	0	1
JUNB	0	1	0	0	0	0	0
KIF18A	0	0	0	1	0	0	0
KIF20A	0	0	0	1	0	0	0
KIF23	0	0	0	1	0	0	0
KIF2A	0	0	0	1	0	0	0
KIF2B	0	0	0	1	0	0	0
KIF2C	0	0	0	1	0	0	0
KL	0	1	0	0	0	0	0
KNTC1	0	0	0	1	0	0	0
KPNA1	1	0	0	0	0	0	0
KPNA2	0	0	0	0	0	0	0
KPNB1	1	0	0	0	0	0	0
KRAS	0	1	0	0	0	0	0
KRT18	1	0	0	0	0	0	0
KRT19	0	1	0	0	0	0	0
KRT8	0	1	0	0	0	0	0
LBP	0	1	0	0	0	0	0
LCK	0	1	0	0	0	0	0
LGALS1	0	0	0	0	0	0	0
LIG1	0	0	1	1	0	1	0
LIG3	0	0	1	0	0	0	0
LIG4	0	0	1	0	0	0	0
LIMK1	1	0	0	0	0	0	0
LMNA	1	0	0	0	0	0	0
LMNB1	1	0	0	0	0	0	0
LMNB2	1	0	0	0	0	0	0
LMO4	0	1	0	0	0	0	0
LRDD	1	1	0	0	1	0	0
LSP1	0	1	0	0	0	0	0

LTA	1	0	0	0	0	0	0
LYN	0	1	0	0	0	0	0
MAD1L1	0	0	0	1	0	0	0
MAD2L1	0	0	0	1	0	0	0
MADD	1	0	0	0	0	0	0
MAGED1	1	0	0	0	0	0	0
MALT1	0	1	0	0	0	0	0
MAP2K1	0	1	0	0	0	0	0
MAP2K2	0	1	0	0	0	0	0
MAP2K3	1	1	0	0	1	0	0
MAP2K4	1	1	0	0	1	0	0
MAP2K6	0	1	0	0	0	0	0
MAP2K7	1	0	0	0	0	0	0
MAP3K1	1	1	0	0	1	0	0
MAP3K10	0	1	0	0	0	0	0
MAP3K11	0	1	0	0	0	0	0
MAP3K12	0	1	0	0	0	0	0
MAP3K13	0	1	0	0	0	0	0
MAP3K14	1	1	0	0	1	0	0
MAP3K2	0	1	0	0	0	0	0
MAP3K3	1	1	0	0	1	0	0
MAP3K4	0	1	0	0	0	0	0
MAP3K5	1	1	0	0	1	0	0
MAP3K6	0	1	0	0	0	0	0
MAP3K7	1	1	0	0	1	0	0
MAP3K7IP2	1	1	0	0	1	0	0
MAP3K8	0	1	0	0	0	0	0
MAP3K9	0	1	0	0	0	0	0
MAP4K2	1	0	0	0	0	0	0
MAP4K3	1	0	0	0	0	0	0
MAP4K4	0	1	0	0	0	0	0
MAP4K5	1	0	0	0	0	0	0
MAPK1	0	1	0	0	0	0	0
MAPK10	0	1	0	0	0	0	0
MAPK11	0	1	0	0	0	0	0
MAPK12	0	1	0	0	0	0	0
MAPK13	0	1	0	0	0	0	0
MAPK14	0	1	0	0	0	0	0
MAPK3	0	1	0	0	0	0	0
MAPK8	1	0	1	0	0	0	0
MAPKAP1	0	1	0	0	0	0	0
MAPKAPK2	0	1	0	0	0	0	0
MAPKAPK3	0	1	0	0	0	0	0
MAPKAPK5	0	0	0	0	0	0	0

MAPRE1	0	0	0	1	0	0	0
MAPT	1	0	0	0	0	0	0
MAX	0	0	0	1	0	0	0
MBD4	0	0	1	0	0	0	0
MCAM	0	0	0	0	0	0	0
MCL1	0	1	0	0	0	0	0
MCM10	0	0	0	1	0	0	0
MCM2	0	0	1	1	0	1	0
MCM3	0	0	0	1	0	0	0
MCM4	0	0	0	1	0	0	0
MCM5	0	0	0	1	0	0	0
MCM6	0	0	0	1	0	0	0
MCM7	0	0	1	1	0	1	0
MCM8	0	0	0	1	0	0	0
MDC1	0	1	1	0	0	0	1
MDM2	0	1	1	1	0	0	1
MEF2A	0	1	0	0	0	0	0
MEF2C	0	1	0	0	0	0	0
MELK	0	0	0	0	0	0	0
METAP2	0	0	0	0	0	0	0
MIR17HG	0	0	0	0	0	0	0
MIS12	0	0	0	1	0	0	0
MITF	0	1	0	0	0	0	0
MKNK1	0	1	0	0	0	0	0
MLF1IP	0	0	0	1	0	0	0
MLH1	0	0	1	0	0	0	0
MLH3	0	0	1	0	0	0	0
MLST8	0	1	0	0	0	0	0
MMS19	0	0	1	0	0	0	0
MNAT1	0	0	1	1	0	1	0
MPG	0	0	1	0	0	0	0
MRE11A	0	1	1	0	0	0	1
MRPL27	0	0	0	0	0	0	0
MSH2	0	0	1	0	0	0	0
MSH3	0	0	1	0	0	0	0
MSH4	0	0	1	0	0	0	0
MSH5	0	0	1	0	0	0	0
MSH6	0	0	1	0	0	0	0
MTHFD1	0	0	0	0	0	0	0
MTOR	0	1	0	0	0	0	0
MUS81	0	0	1	0	0	0	0
MUTYH	0	0	1	0	0	0	0
MXD3	0	0	0	0	0	0	0
MYB	0	0	0	0	0	0	0



MYBL2	0	0	0	0	0	0	0
MYC	0	1	0	1	0	0	1
MYCN	0	0	0	0	0	0	0
MYD88	1	1	0	0	1	0	0
MYT1	0	0	0	1	0	0	0
NBN	0	1	1	0	0	0	1
NCAPD2	0	0	0	0	0	0	0
NDC80	0	0	0	1	0	0	0
NDE1	0	0	0	1	0	0	0
NDEL1	0	0	0	1	0	0	0
NEDD1	0	0	0	1	0	0	0
NEIL1	0	0	1	0	0	0	0
NEIL2	0	0	1	0	0	0	0
NEIL3	0	0	1	0	0	0	0
NEK2	0	0	0	1	0	0	0
NFKB1	1	1	0	0	1	0	0
NFKB2	1	1	0	0	1	0	0
NFKBIA	1	1	1	0	0	0	1
NGF	1	0	0	0	0	0	0
NHEJ1	0	0	1	0	0	0	0
NINL	0	0	0	1	0	0	0
NME1	1	0	0	0	0	0	0
NMT1	1	0	0	0	0	0	0
NOD2	0	1	0	0	0	0	0
NOS2	0	1	0	0	0	0	0
NPAT	0	0	0	0	0	0	0
NR2C2	0	1	0	0	0	0	0
NRAS	0	1	0	0	0	0	0
NRK	1	0	0	0	0	0	0
NRP1	0	0	0	0	0	0	0
NSL1	0	0	0	1	0	0	0
NSMAF	1	0	0	0	0	0	0
NTHL1	0	0	1	0	0	0	0
NTRK1	1	0	0	0	0	0	0
NUDC	0	0	0	1	0	0	0
NUF2	0	0	0	1	0	0	0
NUMA1	1	0	0	1	0	0	1
NUP107	0	0	0	1	0	0	0
NUP133	0	0	0	1	0	0	0
NUP160	0	0	0	1	0	0	0
NUP37	0	0	0	1	0	0	0
NUP43	0	0	0	1	0	0	0
NUP85	0	0	0	1	0	0	0
NUP98	0	0	0	1	0	0	0

OCLN	1	0	0	0	0	0	0
ODF2	0	0	0	1	0	0	0
OFD1	0	0	0	1	0	0	0
OGG1	0	0	1	0	0	0	0
OPN1LW	1	0	0	0	0	0	0
ORC1L	0	0	0	1	0	0	0
ORC2L	0	0	0	1	0	0	0
ORC3L	0	0	0	1	0	0	0
ORC4L	0	0	0	1	0	0	0
ORC5L	0	0	0	1	0	0	0
ORC6L	0	0	0	1	0	0	0
PAFAH1B1	0	0	0	1	0	0	0
PAK1	0	1	0	0	0	0	0
PAK2	1	1	0	0	1	0	0
PAK3	0	1	0	0	0	0	0
PALB2	0	0	1	0	0	0	0
PARP1	1	1	1	0	0	0	1
PARP2	0	0	1	0	0	0	0
PARP3	0	0	1	0	0	0	0
PARP4	0	0	1	0	0	0	0
PBK	0	0	0	0	0	0	0
PCM1	0	0	0	1	0	0	0
PCNA	0	0	1	1	0	1	0
PCNT	0	0	0	1	0	0	0
PDCD4	0	1	0	0	0	0	0
PDPK1	0	1	0	0	0	0	0
PFKFB2	0	0	0	0	0	0	0
PI4KA	0	0	0	1	0	0	0
PIAS1	0	1	0	0	0	0	0
PIAS3	0	1	0	0	0	0	0
PIAS4	0	1	0	0	0	0	0
PIK3CA	1	1	0	0	1	0	0
PIK3CB	1	0	0	0	0	0	0
PIK3CD	1	0	0	0	0	0	0
PIK3CG	1	1	0	0	1	0	0
PIK3R1	1	1	0	0	1	0	0
PIK3R2	1	0	0	0	0	0	0
PIK3R3	1	0	0	0	0	0	0
PIK3R5	1	0	0	0	0	0	0
PIK3R6	0	1	0	0	0	0	0
PKMYT1	0	0	0	1	0	0	0
PKN1	0	0	0	0	0	0	0
PKP1	1	0	0	0	0	0	0
PLA2G4A	0	1	0	0	0	0	0

PLAU	0	0	0	0	0	0	0
PLCB1	0	1	0	0	0	0	0
PLCB2	0	1	0	0	0	0	0
PLCB3	0	1	0	0	0	0	0
PLD1	0	1	0	0	0	0	0
PLD2	0	1	0	0	0	0	0
PLEC1	1	0	0	0	0	0	0
PLK1	0	0	1	1	0	1	0
PLK4	0	0	0	1	0	0	0
PMAIP1	1	0	0	0	0	0	0
PMF1	0	0	0	1	0	0	0
PML	0	1	0	0	0	0	0
PMS1	0	0	1	0	0	0	0
PMS2	0	0	1	0	0	0	0
PMS2L3	0	0	1	0	0	0	0
PMS2L4	0	0	1	0	0	0	0
PNKP	0	0	1	0	0	0	0
PNRC2	0	0	0	0	0	0	0
POLA1	0	0	1	1	0	1	0
POLA2	0	0	0	1	0	0	0
POLB	0	0	0	0	0	0	0
POLD1	0	0	1	1	0	1	0
POLD2	0	0	1	1	0	1	0
POLD3	0	0	1	1	0	1	0
POLD4	0	0	1	1	0	1	0
POLDIP3	0	1	0	0	0	0	0
POLE	0	0	1	1	0	1	0
POLE2	0	0	1	1	0	1	0
POLE3	0	0	1	0	0	0	0
POLE4	0	0	1	0	0	0	0
POLG	0	0	1	0	0	0	0
POLL	0	0	1	0	0	0	0
POLM	0	0	1	0	0	0	0
POT1	0	1	0	0	0	0	0
PPARG	0	0	0	0	0	0	0
PPARGC1A	0	1	0	0	0	0	0
PPM1D	0	1	0	0	0	0	0
PPP1CC	0	0	0	1	0	0	0
PPP1R13B	0	0	0	0	0	0	0
PPP2CA	0	1	1	1	0	0	1
PPP2CB	0	0	0	1	0	0	0
PPP2R1A	0	1	1	1	0	0	1
PPP2R1B	0	0	0	1	0	0	0
PPP2R2B	0	0	1	0	0	0	0

PPP2R5A	0	0	0	1	0	0	0
PPP2R5B	0	0	0	1	0	0	0
PPP2R5C	0	0	0	1	0	0	0
PPP2R5D	0	0	0	1	0	0	0
PPP3CA	1	0	0	0	0	0	0
PPP3CB	1	0	0	0	0	0	0
PPP3CC	1	0	0	0	0	0	0
PPP3R1	1	0	0	0	0	0	0
PPP3R2	1	0	0	0	0	0	0
PRF1	1	0	0	0	0	0	0
PRIM1	0	0	0	1	0	0	0
PRIM2	0	0	0	1	0	0	0
PRKACA	1	0	0	1	0	0	1
PRKACB	1	0	0	0	0	0	0
PRKACG	1	0	0	0	0	0	0
PRKAR1A	1	0	0	0	0	0	0
PRKAR1B	1	0	0	0	0	0	0
PRKAR2A	1	0	0	0	0	0	0
PRKAR2B	1	0	0	1	0	0	1
PRKCA	0	1	0	0	0	0	0
PRKCB	0	1	0	0	0	0	0
PRKCD	1	1	0	0	1	0	0
PRKCE	0	1	0	0	0	0	0
PRKCG	0	1	1	0	0	0	1
PRKCI	1	0	0	0	0	0	0
PRKCQ	1	0	0	0	0	0	0
PRKCZ	1	0	0	0	0	0	0
PRKDC	0	0	1	1	0	1	0
PRR5	0	1	0	0	0	0	0
PSMA1	1	0	0	1	0	0	1
PSMA2	1	0	0	1	0	0	1
PSMA3	1	0	0	1	0	0	1
PSMA4	1	0	0	1	0	0	1
PSMA5	1	0	0	1	0	0	1
PSMA6	1	0	0	1	0	0	1
PSMA7	1	0	0	1	0	0	1
PSMA8	1	0	0	1	0	0	1
PSMB1	1	0	0	1	0	0	1
PSMB10	1	0	0	1	0	0	1
PSMB11	1	0	0	1	0	0	1
PSMB2	1	0	0	1	0	0	1
PSMB3	1	0	0	1	0	0	1
PSMB4	1	0	0	1	0	0	1
PSMB5	1	0	0	1	0	0	1

PSMB6	1	0	0	1	0	0	1
PSMB7	1	0	0	1	0	0	1
PSMB8	1	0	0	1	0	0	1
PSMB9	1	0	0	1	0	0	1
PSMC1	1	0	0	1	0	0	1
PSMC2	1	0	0	1	0	0	1
PSMC3	1	0	0	1	0	0	1
PSMC4	1	0	0	1	0	0	1
PSMC5	1	0	0	1	0	0	1
PSMC6	1	0	0	1	0	0	1
PSMD1	1	0	0	1	0	0	1
PSMD10	1	0	0	1	0	0	1
PSMD11	1	0	0	1	0	0	1
PSMD12	1	0	0	1	0	0	1
PSMD13	1	0	0	1	0	0	1
PSMD14	1	0	0	1	0	0	1
PSMD2	1	0	0	1	0	0	1
PSMD3	1	0	0	1	0	0	1
PSMD4	1	0	0	1	0	0	1
PSMD5	1	0	0	1	0	0	1
PSMD6	1	0	0	1	0	0	1
PSMD7	1	0	0	1	0	0	1
PSMD8	1	0	0	1	0	0	1
PSMD9	1	0	0	1	0	0	1
PSME1	1	0	0	1	0	0	1
PSME2	1	0	0	1	0	0	1
PSME3	1	0	0	1	0	0	1
PSME4	1	0	0	1	0	0	1
PSMF1	1	0	0	1	0	0	1
PTGER2	0	1	0	0	0	0	0
PTGES2	0	1	0	0	0	0	0
PTGS2	0	1	0	0	0	0	0
PTK2	1	0	0	0	0	0	0
PTPN11	0	1	0	0	0	0	0
PTPRE	0	1	0	0	0	0	0
PTTG1	0	0	0	1	0	0	0
PURA	0	0	0	0	0	0	0
PXN	0	1	0	0	0	0	0
RAB11A	0	1	0	0	0	0	0
RAB5A	0	1	0	0	0	0	0
RAB7A	0	1	0	0	0	0	0
RAC1	0	1	0	0	0	0	0
RAC2	0	1	0	0	0	0	0
RAD1	0	0	1	1	0	1	0

RAD17	0	0	1	1	0	1	0
RAD21	0	0	0	1	0	0	0
RAD23A	0	0	1	0	0	0	0
RAD23B	0	0	1	0	0	0	0
RAD50	0	1	1	0	0	0	1
RAD51	0	0	1	0	0	0	0
RAD51AP1	0	0	0	0	0	0	0
RAD51C	0	0	1	0	0	0	0
RAD51L1	0	0	1	0	0	0	0
RAD51L3	0	0	1	0	0	0	0
RAD52	0	0	1	0	0	0	0
RAD54B	0	0	1	0	0	0	0
RAD54L	0	0	1	0	0	0	0
RAD9A	0	0	1	1	0	1	0
RAD9B	0	0	0	1	0	0	0
RAF1	0	1	0	0	0	0	0
RAN	0	1	0	0	0	0	0
RANBP2	0	0	0	1	0	0	0
RANGAP1	0	0	0	1	0	0	0
RASGRF1	0	0	0	1	0	0	0
RB1	0	1	0	1	0	0	1
RB1CC1	0	1	0	0	0	0	0
RBBP4	0	0	0	0	0	0	0
RBBP8	0	0	1	0	0	0	0
RBL1	0	0	0	1	0	0	0
RBM25	0	0	0	0	0	0	0
RBX1	0	0	1	0	0	0	0
RCC2	0	0	0	1	0	0	0
REC8	0	0	0	1	0	0	0
REL	0	1	0	0	0	0	0
RELA	1	1	1	0	0	0	1
RELB	0	1	0	0	0	0	0
RFC1	0	0	1	1	0	1	0
RFC2	0	0	1	1	0	1	0
RFC3	0	0	1	1	0	1	0
RFC4	0	0	1	1	0	1	0
RFC5	0	0	1	1	0	1	0
RFFL	1	0	0	0	0	0	0
RFWD2	0	0	0	1	0	0	0
RHEB	0	1	0	0	0	0	0
RHOA	0	1	0	0	0	0	0
RICTOR	0	1	0	0	0	0	0
RIPK1	1	1	0	0	1	0	0
RIPK2	0	1	0	0	0	0	0

RNF168	0	0	0	0	0	0	0
RNF8	0	1	0	0	0	0	0
ROCK1	1	0	0	0	0	0	0
RPA1	0	1	1	1	0	0	1
RPA2	0	1	1	1	0	0	1
RPA3	0	1	1	1	0	0	1
RPA4	0	1	1	1	0	0	1
RPL10	0	0	0	0	0	0	0
RPL10A	0	0	0	0	0	0	0
RPL11	0	0	0	0	0	0	0
RPL12	0	0	0	0	0	0	0
RPL13	0	0	0	0	0	0	0
RPL13A	0	0	0	0	0	0	0
RPL14	0	0	0	0	0	0	0
RPL15	0	0	0	0	0	0	0
RPL17	0	0	0	0	0	0	0
RPL18	0	0	0	0	0	0	0
RPL18A	0	0	0	0	0	0	0
RPL19	0	0	0	0	0	0	0
RPL21	0	0	0	0	0	0	0
RPL22	0	0	0	0	0	0	0
RPL23	0	0	0	0	0	0	0
RPL23A	0	0	0	0	0	0	0
RPL24	0	0	0	0	0	0	0
RPL26	0	0	0	0	0	0	0
RPL26L1	0	0	0	0	0	0	0
RPL27	0	0	0	0	0	0	0
RPL27A	0	0	0	0	0	0	0
RPL28	0	0	0	0	0	0	0
RPL29	0	0	0	0	0	0	0
RPL3	0	0	0	0	0	0	0
RPL30	0	0	0	0	0	0	0
RPL31	0	0	0	0	0	0	0
RPL32	0	0	0	0	0	0	0
RPL34	0	0	0	0	0	0	0
RPL35	0	0	0	0	0	0	0
RPL35A	0	0	0	0	0	0	0
RPL36	0	0	0	0	0	0	0
RPL36A	0	0	0	0	0	0	0
RPL37	0	0	0	0	0	0	0
RPL37A	0	0	0	0	0	0	0
RPL38	0	0	0	0	0	0	0
RPL39	0	0	0	0	0	0	0
RPL3L	0	0	0	0	0	0	0

RPL4	0	0	0	0	0	0	0
RPL41	0	0	0	0	0	0	0
RPL5	0	0	0	0	0	0	0
RPL6	0	0	0	0	0	0	0
RPL7	0	0	0	0	0	0	0
RPL7A	0	0	0	0	0	0	0
RPL8	0	0	0	0	0	0	0
RPL9	0	0	0	0	0	0	0
RPL9P7	0	0	0	0	0	0	0
RPL9P8	0	0	0	0	0	0	0
RPL9P9	0	0	0	0	0	0	0
RPLP0	0	0	0	0	0	0	0
RPLP1	0	0	0	0	0	0	0
RPLP2	0	0	0	0	0	0	0
RPS10	0	0	0	0	0	0	0
RPS11	0	0	0	0	0	0	0
RPS12	0	0	0	0	0	0	0
RPS13	0	0	0	0	0	0	0
RPS14	0	0	0	0	0	0	0
RPS15	0	0	0	0	0	0	0
RPS15A	0	0	0	0	0	0	0
RPS16	0	0	0	0	0	0	0
RPS17	0	0	0	0	0	0	0
RPS18	0	0	0	0	0	0	0
RPS19	0	0	0	0	0	0	0
RPS2	0	0	0	0	0	0	0
RPS20	0	0	0	0	0	0	0
RPS21	0	0	0	0	0	0	0
RPS23	0	0	0	0	0	0	0
RPS24	0	0	0	0	0	0	0
RPS25	0	0	0	0	0	0	0
RPS26	0	0	0	0	0	0	0
RPS27	0	0	0	1	0	0	0
RPS27A	1	0	1	1	0	0	1
RPS28	0	0	0	0	0	0	0
RPS29	0	0	0	0	0	0	0
RPS3	0	0	0	0	0	0	0
RPS3A	0	0	0	0	0	0	0
RPS4X	0	0	0	0	0	0	0
RPS4Y1	0	0	0	0	0	0	0
RPS5	0	0	0	0	0	0	0
RPS6	0	0	0	0	0	0	0
RPS6KA1	0	1	0	1	0	0	1
RPS6KA4	0	1	0	0	0	0	0



RPS6KA5	0	1	0	0	0	0	0
RPS6KB1	0	1	0	0	0	0	0
RPS7	0	0	0	0	0	0	0
RPS8	0	0	0	0	0	0	0
RPS9	0	0	0	0	0	0	0
RPSA	0	0	0	0	0	0	0
RPTOR	0	1	0	0	0	0	0
RQCD1	0	0	0	0	0	0	0
RRAGA	0	1	0	0	0	0	0
RRAGB	0	1	0	0	0	0	0
RRAGC	0	1	0	0	0	0	0
RRAGD	0	1	0	0	0	0	0
RRM1	0	0	0	0	0	0	0
RRM2	0	0	0	1	0	0	0
RRN3	0	1	0	0	0	0	0
SATB1	1	0	0	0	0	0	0
SDCCAG8	0	0	0	1	0	0	0
SEC13	0	0	0	1	0	0	0
SEH1L	0	0	0	1	0	0	0
SERPINE1	0	0	0	0	0	0	0
SET	1	0	0	0	0	0	0
SFI1	0	0	0	1	0	0	0
SFN	0	1	0	1	0	0	1
SGK1	0	1	0	0	0	0	0
SGOL1	0	0	0	1	0	0	0
SGOL2	0	0	0	1	0	0	0
SHFM1	0	0	1	0	0	0	0
SIVA1	0	0	0	0	0	0	0
SKA1	0	0	0	1	0	0	0
SKA2	0	0	0	1	0	0	0
SKP1	0	0	0	1	0	0	0
SKP2	0	0	0	1	0	0	0
SLC25A14	0	0	0	0	0	0	0
SLC3A2	0	0	0	0	0	0	0
SLC9A1	0	1	0	0	0	0	0
SLK	1	0	0	0	0	0	0
SMAD3	0	0	0	0	0	0	0
SMAD4	0	0	0	0	0	0	0
SMARCAL1	0	0	1	0	0	0	0
SMC1A	0	0	0	1	0	0	0
SMC3	0	0	0	1	0	0	0
SMPD1	1	0	0	0	0	0	0
SMPD2	1	0	0	0	0	0	0
SMUG1	0	0	1	0	0	0	0

SNRPE	0	0	0	0	0	0	0
SNTA1	0	0	0	0	0	0	0
SNX19	0	0	0	0	0	0	0
SOAT2	0	0	0	0	0	0	0
SOCS3	0	1	0	0	0	0	0
SOS1	0	1	0	0	0	0	0
SP1	0	1	0	0	0	0	0
SPAG7	0	0	0	0	0	0	0
SPC24	0	0	0	1	0	0	0
SPC25	0	0	0	1	0	0	0
SPTAN1	1	0	0	0	0	0	0
SQSTM1	1	0	0	0	0	0	0
SRC	0	1	0	0	0	0	0
SREBF1	1	1	0	0	1	0	0
SRF	0	1	0	0	0	0	0
SRPR	0	0	0	0	0	0	0
SSBP1	0	0	1	0	0	0	0
SSNA1	0	0	0	1	0	0	0
SSPO	0	1	1	0	0	0	1
SSR4	0	0	0	0	0	0	0
STAG1	0	0	0	1	0	0	0
STAG2	0	0	0	1	0	0	0
STAT1	1	1	0	0	1	0	0
STAT2	0	1	0	0	0	0	0
STAT3	0	1	0	0	0	0	0
STAT4	0	1	0	0	0	0	0
STAT5A	0	1	0	0	0	0	0
STAT5B	0	1	0	0	0	0	0
STAT6	0	1	0	0	0	0	0
STK24	1	0	0	0	0	0	0
STMN1	0	0	0	0	0	0	0
SUZ12	0	0	0	0	0	0	0
SYK	0	1	0	0	0	0	0
TAB1	1	1	0	0	1	0	0
TANK	1	0	0	0	0	0	0
TAOK1	0	1	0	1	0	0	1
TAOK2	0	1	0	0	0	0	0
TAOK3	0	1	0	0	0	0	0
TCF3	0	1	0	0	0	0	0
TDG	0	0	1	0	0	0	0
TDP1	0	0	1	0	0	0	0
TERF2	0	1	0	0	0	0	0
TERF2IP	0	1	0	0	0	0	0
TERT	0	1	0	0	0	0	0

TFAP2A	1	0	0	0	0	0	0
TFDP1	1	0	0	1	0	0	1
TGFB1	0	1	0	1	0	0	1
TH	0	1	0	0	0	0	0
TIMELESS	0	0	1	0	0	0	0
TIMP1	0	1	0	0	0	0	0
TIPIN	0	0	1	0	0	0	0
TJP1	1	0	0	0	0	0	0
TJP2	1	0	0	0	0	0	0
TK1	0	0	0	0	0	0	0
TK2	0	0	0	1	0	0	0
TLR4	0	1	0	0	0	0	0
TNF	1	1	0	0	1	0	0
TNFAIP3	1	1	0	0	1	0	0
TNFRSF10A	1	0	0	0	0	0	0
TNFRSF10B	1	0	0	0	0	0	0
TNFRSF10C	1	0	0	0	0	0	0
TNFRSF10D	1	0	0	0	0	0	0
TNFRSF1A	1	1	0	0	1	0	0
TNFRSF1B	1	0	0	0	0	0	0
TNFRSF25	1	0	0	0	0	0	0
TNFSF10	1	0	0	0	0	0	0
TNFSF11	0	1	0	0	0	0	0
TNFSF12	1	0	0	0	0	0	0
TNIK	1	0	0	0	0	0	0
TOP1	1	0	0	0	0	0	0
TOP2A	0	0	0	0	0	0	0
TOP3A	0	0	1	0	0	0	0
TOP3B	0	0	1	0	0	0	0
TOPBP1	0	1	1	0	0	0	1
TP53	1	1	1	1	0	0	1
TP53BP1	0	1	1	0	0	0	1
TP53BP2	0	0	0	0	0	0	0
TP53INP1	0	0	0	0	0	0	0
TP73	0	0	1	0	0	0	0
TRADD	1	1	0	0	1	0	0
TRAF1	1	0	0	0	0	0	0
TRAF2	1	0	0	0	0	0	0
TRAF3	1	0	0	0	0	0	0
TRAF6	0	1	0	0	0	0	0
TREX1	0	0	1	0	0	0	0
TRIM55	0	0	0	0	0	0	0
TSC1	0	1	0	0	0	0	0
TSC2	0	1	0	0	0	0	0

TSGA14	0	0	0	1	0	0	0
TUBA1A	0	0	0	1	0	0	0
TUBA4A	0	0	0	1	0	0	0
TUBB	0	0	0	1	0	0	0
TUBB2C	0	0	0	1	0	0	0
TUBB4	0	0	0	1	0	0	0
TUBG1	0	0	0	1	0	0	0
TUBG2	0	0	0	1	0	0	0
TUBGCP2	0	0	0	1	0	0	0
TUBGCP3	0	0	0	1	0	0	0
TUBGCP4	0	0	0	1	0	0	0
TUBGCP5	0	0	0	1	0	0	0
TUBGCP6	0	0	0	1	0	0	0
TXN	1	0	0	0	0	0	0
TYK2	0	1	0	0	0	0	0
TYMS	0	0	0	1	0	0	0
UBA52	1	0	1	1	0	0	1
UBB	1	0	1	1	0	1	1
UBC	1	0	1	1	0	1	1
UBE2C	0	0	0	1	0	0	0
UBE2D1	0	0	0	1	0	0	0
UBE2D3	0	1	0	0	0	0	0
UBE2E1	0	0	0	1	0	0	0
UBE2N	0	1	1	0	0	0	1
UBE2T	0	0	1	0	0	0	0
UHRF1	0	0	0	0	0	0	0
UIMC1	0	1	0	0	0	0	0
ULK1	0	1	0	0	0	0	0
ULK2	0	1	0	0	0	0	0
UNC5A	1	0	0	0	0	0	0
UNC5B	1	0	0	0	0	0	0
UNG	0	0	1	0	0	0	0
USF1	0	1	0	0	0	0	0
USP1	0	0	1	0	0	0	0
UXT	0	0	0	0	0	0	0
VASP	0	0	0	0	0	0	0
VAV1	0	1	0	0	0	0	0
VIM	1	0	0	0	0	0	0
WDR36	0	0	0	0	0	0	0
WEE1	0	0	0	1	0	0	0
WNT16	0	0	0	0	0	0	0
WNT2	0	1	0	0	0	0	0
WRN	0	0	1	0	0	0	0
WWOX	0	0	0	0	0	0	0

XAB2	0	0	1	0	0	0	0
XIAP	1	1	0	0	1	0	0
XPA	0	0	1	0	0	0	0
XPC	0	0	1	0	0	0	0
XPO1	0	1	0	1	0	0	1
XRCC1	0	0	1	0	0	0	0
XRCC2	0	0	1	0	0	0	0
XRCC3	0	0	1	0	0	0	0
XRCC4	0	0	1	0	0	0	0
XRCC5	0	0	1	0	0	0	0
XRCC6	0	0	1	0	0	0	0
YES1	0	1	0	0	0	0	0
YWHAB	1	1	1	0	0	0	1
YWHAE	0	1	0	1	0	0	1
YWHAG	0	1	0	1	0	0	1
YWHAH	0	1	0	1	0	0	1
YWHAQ	0	1	0	0	0	0	0
YWHAZ	0	1	1	0	0	0	1
YY1	0	1	0	0	0	0	0
ZBTB32	0	0	1	0	0	0	0
ZNF451	0	0	0	0	0	0	0
ZNF83	0	0	0	0	0	0	0
ZW10	0	0	0	1	0	0	0
ZWILCH	0	0	0	1	0	0	0
ZWINT	0	0	0	1	0	0	0